University of Alberta

Efficient Detection Techniques for Wireless Communication Systems

by

Tao Cui

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Electrical and Computer Engineering

Edmonton, Alberta

Fall 2005

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-09148-7

Your file *Votre référence*
ISBN:
Our file *Notre reference*
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

For my parents, who offered me endless encouragement, love and support.

Abstract

Efficient receiver design plays a crucial role in designing next-generation wireless communication systems for high data-rate transmission. In this thesis, we propose a unified framework for tree search based data detectors for multiple-input and multiple-output systems. We also develop constrained linear detectors, decision feedback detectors and multistage detectors.

Blind data detectors for orthogonal space-time block code are derived. We also develop an efficient detector for differential unitary space-time modulation (DUSTM), which is termed as bound-intersection detector (BID). The extended Euclidean algorithm is used to generate the candidate sets. Our BID achieves significant computational savings over the exhaustive maximum likelihood search. Moreover, we develop BID variants for multiple symbol detection of DUSTM over both Rayleigh and Ricean channels. Furthermore, blind and semi-blind data detectors for orthogonal frequency division multiplexing are derived. Channel parameter estimation algorithms are also derived.

Acknowledgements

This is perhaps the easiest and hardest chapter that I have to write. It will be simple to name all the people who made this thesis possible, but it will be tough to thank them enough.

My greatest debt is to my supervisor Dr. Chintha Tellambura for his valuable advice and constant encouragement. His insight, kindness and genuine concern for his students made working with him a memorable experience. I could not have imagined having a better advisor and mentor for my master degree. I am grateful to him for leading me into this exciting area of wireless communications.

Thank-you to my examiners, Dr. Masoud Ardakani and Dr. Mike MacGregor, for serving on my thesis committee and managing to read the whole thing so thoroughly. I would like to thank Ms. Sandra Abello and all the support staffs of the Informatics Circle of Research Excellence (iCORE) Wireless Communications Laboratory.

Much respect to my current labmates, and hopefully still friends. Thank Yu Fu, Luqing Wang and Yue Wu for providing me with invaluable help and advice for my life here. I am grateful to Dr. Wen Chen, Saeed Fouladi Fard and Alireza Ghaderipoor for their helpful discussion. My gratitude also goes to other labmates for providing such a creative, intellectually stimulating and fun environment.

I am forever indebted to my parents for their understanding, endless love, patience and encouragement when it was most required. I feel fortunate to have been brought up by two of the most kind people I have ever met.

Last, I would like to thank the Alberta Ingenuity Fund and iCORE for their financial support.

Contents

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Examples of Detection Algorithms 4

1.2.1 V-BLAST algorithm 6

1.2.2 Sphere decoding algorithm 8

1.3 Thesis Contributions 10

2 Efficient Data Detection for Uncoded MIMO Systems **14**

2.1 Introduction 14

2.1.1 Background 14

2.1.2 System model 17

2.2 Unified Framework for Tree Search Based Detection 20

2.2.1 Feedback decoding 20

2.2.2 New tree interpretation of MIMO 23

2.2.3 Computation sharing technique 26

2.2.4 Performance analysis 28

2.2.5 Computational complexity analysis 32

2.2.6 Simulation results 34

2.3 Constrained Detection for MIMO Systems 40

2.3.1 Classic linear detectors 40

2.3.2 Real constrained detectors 41

2.3.3 Constrained subgroup detectors based on constant-modulus constellations 41

2.3.4 Chase improvement 44
2.3.5 Real decision feedback detectors 44
2.3.6 Constrained ordering decision feedback detectors 45
2.3.7 Combined constrained linear and decision feedback detectors . 46
2.3.8 Polynomial constrained detector 47
2.3.9 Simulation results 50
2.4 Multistage Sphere Decoder for High Order Constellations 56
2.4.1 Multistage sphere decoding algorithm 56
2.4.2 Simulation results 59
2.5 Conclusion 61

3 Blind Decoding for Orthogonal Space-Time Block Codes 63
3.1 Introduction 63
3.1.1 Background 63
3.1.2 System model 65
3.2 Maximum-Likelihood Blind Decoder 66
3.3 Totally Blind Decoders 71
3.3.1 Different constellations scheme 71
3.3.2 Superimposed pilots scheme 72
3.4 Channel Estimation 73
3.5 OSTBC Decoding over Time-Selective Fading Channels with Perfect CSI 74
3.6 Simulation Results 75
3.7 Conclusion 77

4 Efficient Detection of Multiple-Symbol Differential Unitary Space-Time Modulation 79
4.1 Introduction 79
4.1.1 Background 79
4.1.2 System model and differential unitary space-time modulation . 82

4.2 Reduced Complexity Differential Unitary Space-Time Demodulation over Rayleigh MIMO Channels 84
4.2.1 Decision metric 84
4.2.2 Reduced complexity single-symbol detection 86
4.2.3 Reduced complexity multiple-symbol detection 91
4.3 Reduced Complexity Differential Unitary Space-Time Demodulation over Ricean MIMO Channels 96
4.3.1 Decision metric 96
4.3.2 Reduced complexity multiple-symbol detection 98
4.4 Simulation Results 101
4.4.1 Rayleigh channels 101
4.4.2 Ricean channels 109
4.5 Conclusion 114

5 Blind and Semi-Blind Data Detection for OFDM Systems 115
5.1 Introduction 116
5.1.1 Background 116
5.1.2 OFDM baseband model 117
5.2 Blind and Semiblind Data Detectors 119
5.2.1 Blind detector 120
5.2.2 Semi-blind detector 122
5.2.3 Enhanced data detector with channel estimation error 123
5.3 Data Detection Algorithms 125
5.4 Power Delay Profile and Noise Variance Estimation 127
5.5 Generalizations 132
5.5.1 Semi-blind detector under mismatch 132
5.5.2 Generalization to non-unitary constellations 134
5.5.3 Time-varying channel tracking via decision feedback 136
5.6 Simulation Results 136
5.7 Conclusion 141

6 Conclusion 144

References 147

List of Tables

2.1 Relationship to the generalized feedback detector. 24

4.1 Complexity comparison for ML, LLL and BID in flops. 103

List of Figures

2.1 Basic algorithm illustration of the GFD with different w_k and s_k. . . 22

2.2 Comparison of exhaustive search, GFD and ZF-DFD in tree representation for $N_T = N_R = 4$ and BPSK. 27

2.3 BER comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$. 34

2.4 Average complexity comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$. 35

2.5 BER comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$. 36

2.6 Average complexity comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$. 36

2.7 BER comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$. 37

2.8 Average complexity comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$. 38

2.9 BER comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$. 39

2.10 Average complexity comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$. 39

2.11 Performance comparison of constrained linear detectors in an 8×8 MIMO system with BPSK. 51

2.12 Complexity of constrained linear detectors in an 8×8 MIMO system with 16QAM. 52

2.13 Complexity of constrained ordering decision feedback detectors in an 8×8 MIMO system with BPSK. 52
2.14 Complexity of combined linear and decision feedback detectors in an 8×8 BPSK MIMO system. 53
2.15 The trajectory of a particle for a 2×2 MIMO system with BPSK and 5 dB. 54
2.16 The energy of a particle as a function of time steps for a 2×2 MIMO system with BPSK and 5 dB. 55
2.17 Performance comparison of different detectors in an 8×8 MIMO system with BPSK. 56
2.18 Performance comparison of different detectors in an 4×4 MIMO system with 16QAM. 57
2.19 The block diagram of the multistage sphere decoder for 16QAM systems. 57
2.20 BER comparison with different N_{cand} for a 16QAM MIMO system with 4 transmit and 4 receive antennas. 61
2.21 Complexity comparison with different N_{cand} for a 16QAM MIMO system with 4 transmit and 4 receive antennas. 62

3.1 The transmission diagram of a space time block coded system. 67
3.2 BER versus SNR for different blind decoders with $N = 8$ and BPSK. 76
3.3 Average flops versus SNR for different blind decoders with $N = 8$ and BPSK. 77
3.4 BER versus SNR for different blind decoders with $N = 8$, BPSK and $f_d T_B = 0.005$. 78

4.1 Performance comparison for $N_T = 3, 4, 5$ transmitter antennas, $N_R = 1$ receiver antenna as a function of SNR. The channel is static fading and $R = 2$. 102
4.2 Complexity of BID for $N_T = 3, 4, 5$ transmitter antennas, $N_R = 1$ receiver antenna versus SNR. The channel is static fading and $R = 2$. 103

4.3 Complexity comparison between BID and LLL for $N_T = 4$ transmitter antennas, $N_R = 1, 2, 3, 4$ receiver antenna versus SNR. The channel is static fading and $R = 2$. 105
4.4 Performance comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ and $R = 1$ as a function of SNR. The channel is constant within N blocks. 106
4.5 Complexity comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ and $R = 1$ as a function of SNR. The channel is constant within N blocks. 106
4.6 Performance comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ and $R = 1$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.0075$ and $R = 1$. 107
4.7 Complexity comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.0075$ and $R = 1$. 108
4.8 Performance comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.03$ and $R = 1$. 108
4.9 Complexity comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.03$ and $R = 1$. 109
4.10 Complexity of MSD2 for $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with SNR=30 dB, $R = 1$ and different normalized Doppler frequencies as a function of N. 110
4.11 The performance comparison between SD-BID, MSD based DF-DD, CDD and CD with $N = 3, 6$ for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.0075$ and $K = 5$ dB). 111
4.12 The performance comparison between SD-BID, MSD based DF-DD, CDD and CD with $N = 3, 6$ for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$ and $K = 5$ dB). 111

4.13 The performance comparison between SD-BID ($N = 6$), CDD and CD for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$) with different Rice factor K. 112
4.14 The average number of flops comparison between SD-BID, RS-DD, MSD based DF-DD for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$ and $K = 5$ dB). 113
4.15 The average number of flops of SD-BID for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels for different $f_D T_B$ and Rice factor K. 113

5.1 The probability of correct detection of the number of paths. 131
5.2 The NMSE of the channel power estimation for the 3rd path. 132
5.3 The NMSE of the noise variance estimation using different estimators. 133
5.4 MSE of the joint ML estimation of the channel response versus SNR for an OFDM system with $N = 32$ and BPSK in a 6-ray exponential PDP channel. 137
5.5 Bit error rate of the joint ML estimation algorithm versus SNR for an OFDM system with $N = 32$ and BPSK in a 6-ray exponential PDP channel. 138
5.6 The computational complexity versus SNR for an OFDM system with $N = 32$ and BPSK in a 6-ray exponential PDP channel. 139
5.7 Bit error rate of the joint ML estimation algorithm versus SNR for an OFDM system with $N = 32$ and BPSK in a TU channel 140
5.8 Effects of semi-blind detector design mismatch in an OFDM system with $N = 32$ and BPSK. The channel is simulated using an exponential PDP. But in the semi-blind detector, the uniform PDP is assumed . . 141
5.9 Bit error rate of the joint ML estimation algorithm versus SNR for an OFDM system with $N = 32$ and 4PAM in a 3-ray exponential PDP channel. 142
5.10 The BER versus SNR for an BPSK OFDM system with $N = 32$ and assuming different channel length L. 143

5.11 The performance of the enhanced data detector as a function of the number of pilots for a BPSK-OFDM system with $N = 32$. 143

Acronyms

Acronyms	Definition
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BnB	Branch-and-Bound
BPSK	Binary Phase Shift Keying
CDMA	Code Division Multiple Access
CGRV	Complex Gaussian Random Variable
CIR	Channel Impulse Response
CP	Cyclic Prefix
CRB	Cramer-Rao Bound
CSI	Channel State Information
CVP	Closest Vector Problem
DFD	Decision Feedback Detector
DFT	Discrete Fourier Transform
DPSK	Differential Phase-Shift Keying
DUSTM	Differential Unitary Space-Time Modulation
FDMA	Frequency Division Multiple Access
gcd	greatest common divisor
LS	Least Squares
MIMO	Multiple-Input and Multiple-Output
ML	Maximum-Likelihood
MMSE	Minimum Mean Squared Error
MSE	Mean Square Error

OFDM	Orthogonal Frequency Division Multiplexing
pdf	probability density function
PAM	Pulse Amplitude Modulation
PDP	Power Delay Profile
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
QS	Quasi Static
SD	Sphere Decoder
SER	Symbol Error Rate
SINR	Signal-to-Interference and Noise Ratio
SNR	Signal-to-Noise Ratio
STBC	Space-Time Block Code
TDMA	Time Division Multiple Access
V-BLAST	Vertical Bell Laboratories Layered Space-Time Architecture
VLSI	Very Large Scale Integration
ZF	Zero Forcing

Chapter 1

Introduction

1.1 Background and Motivation

Driven by the demand for increasingly sophisticated connectivity anytime, anywhere, wireless communications has emerged as one of the largest and most rapidly growing sectors of the global telecommunications industry. From broadcast radio stations to cellular telephones to wireless Internet, there are more wireless applications than ever before.

The first generation systems were introduced in the mid 1980's, and can be characterized by the use of analog transmission techniques, and the use of simple multiple access techniques such as frequency division multiple access (FDMA). Those systems such as Advanced Mobile Phone Service (AMPS) only provided voice communications. They also suffered from a low user capacity, and security problems due to the simple radio interface.

Second generation systems were introduced in the early 1990's, and all used digital technology. This increases the user capacity almost three times. This increase was achieved by compressing the voice waveforms before transmission.

Third generation systems are an extension of second-generation systems, and operators have begun roll out of services since 2001. The capacity of third generation systems is over ten times that of original first generation systems. This high capacity is achieved by using complex multiple access techniques such as code division multiple

access (CDMA), or an extension of time division multiple access (TDMA), and by improving the flexibility of services available.

However, the consumers' demand for wireless multimedia services and high-speed Internet access has been growing. As a result, the wireless industry continues to evolve, and the demand for more sophisticated communications technologies grows. Multiple-input and multiple-output (MIMO) and orthogonal frequency division multiplexing (OFDM) are two technologies under consideration for the future fourth generation (4G) wireless networks providing multimedia services.

MIMO is a communications technique that uses multiple antennas to send and receive wireless signals, allowing more data to be transmitted without increasing bandwidth. This is accomplished by communicating along parallel spatial channels at the same time and in the same frequency. It has been both shown theoretically and demonstrated in experimental laboratory settings that MIMO systems over a rich scattering wireless channel are capable of providing enormous capacity improvements without increasing the bandwidth or transmitted power. Hence, the recent explosion of interest is from both academic and industrial researchers in the area of signal processing techniques for MIMO systems.

OFDM is a multicarrier transmission technique, which divides the available spectrum into many sub-carriers. Each one is modulated by a low data rate stream. OFDM is similar to FDMA in that multiple access is achieved by subdividing the available bandwidth into multiple channels, which are then allocated to users. However, OFDM uses the spectrum much more efficiently by spacing the channels much closer together. This is achieved by making all sub-carriers orthogonal to one another, preventing interference between the closely spaced carriers.

Many standards have been developed for further global wireless systems, with more standards likely to emerge. The performance improvements resulting from new standards usually come at the cost of increased computational complexity in the receiver (and often the transmitter as well). The design of low-complexity receivers is, therefore, one of the key problems in 4G wireless system design. The largest potential for complexity reduction of highest-performance very large scale integration

(VLSI) signal processing circuit depends highly on the optimization of algorithms. In this thesis, we optimize detection algorithms for MIMO and OFDM systems, which is critical to achieve practicable solutions for 4G wireless systems.

MIMO signal detection has been the subject of intensive study. A prime example is the Bell laboratories layered space time (BLAST) architecture [1], and the related V-BLAST detector, which is a suboptimal detector. Recently, much effort has been expended on the sphere decoder (SD) [2], which is considered to be one of the most important and industrially relevant algorithms. Maximum likelihood (ML) or optimal detection of signals transmitted over MIMO channels is well-known to be an NP-hard problem. However, the SD has been shown to offer ML detection at a computational complexity that is polynomial in the average case [3], where the number of antennas is moderate and the signal-to-noise ratio (SNR) is high. Hardware implementations of the SD have already been reported in the literature [4,5]. Even so, existing SDs exhibit two major weaknesses. First, the performance of most current SD algorithms is highly sensitive to the value chosen for the initial search radius. The successful termination of the algorithm, i.e., returning an optimal solution, as well as its time complexity, is also heavily dependent on the initial search radius [6]. Secondly, although its time complexity is polynomial in the average case, the complexity can become very large when the SNR is low, or when the problem dimension is high, e.g., at the high spectral efficiencies required to support higher communication rates [6]. These motivate the development of near ML detectors that have affordable worst-case complexity in all the cases. We will present these detectors in Chapter 2.

Space-time coding techniques for MIMO system have recently emerged as a promising way for effectively utilizing the multi-antenna diversity. In particular, space-time block codes (STBCs) based on the theory of orthogonal designs [7] have attracted much attention because they achieve the maximum antenna diversity gain with a simple code structure. In addition, ML decoding of orthogonal STBCs (OSTBC) requires only simple linear processing, if the channel fading coefficients are known. However, a multiple antenna channel is difficult to estimate and may vary rapidly due to the users' mobility. Moreover, channel state information (CSI) estimation using

pilot symbols reduces the effective data rate. These factors motivates blind OSTBC detection. In Chapter 3, we investigate ML blind detection of OSTBCs.

Differential space-time modulation (DSTM) [8–10], particularly differential unitary space-time modulation (DUSTM), allows MIMO communication entirely without the possession of CSI by either the transmitter or the receiver. Since DUSTM generalizes the classical single-antenna differential phase-shift keying (DPSK), similar to DPSK, DUSTM performs 3 dB worse than its coherent counterpart. To improve the performance of single symbol differential detection, attempts have been made to extend some of these detection techniques to the multiple-antenna case. In Chapter 4, efficient algorithms for multiple symbol detection of DUSTM are developed.

In OFDM systems, coherent data detection also requires CSI. The use of a cyclic prefix (CP) and pilot tones for channel estimation constitutes a significant overhead or bandwidth loss, motivating the development of blind techniques for OFDM. Several blind channel estimators have been proposed by using statistical or deterministic properties of the transmit and receive signals. However, most of these blind estimators typically use averaging over a large number of OFDM symbols (up to several thousands in some cases). These estimators thereby introduce a considerable latency into the overall system and have high complexity. We will develop several blind data detectors for OFDM systems requiring only one OFDM symbol and give their efficient implementations in Chapter 5.

1.2 Examples of Detection Algorithms

Many detection problems considered in this thesis may be based on the system model formulated in a lattice as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1.1}$$

where $\mathbf{x} \in \mathcal{Z}^n$, $\mathbf{r}, \mathbf{n} \in \mathcal{R}^m$ denote the system input, output and additive noise, and $\mathbf{H} \in \mathcal{R}^{m \times n}$ represents the transfer matrix (i.e., the channel matrix in MIMO system), $\mathcal{Z}^n$ denotes the n-dimensional integer lattice. Generally, the noise terms n_i, $i = 1, \cdots, m$ are independent and identically distributed (i.i.d.) zero-mean Gaussian

random variables with the same variance σ_n^2. In this chapter, we assume that $n \geq m$ and $\mathbf{H}$ has full column rank. Under such conditions and assuming $\mathbf{H}$ is perfectly known at the receiver, the optimal ML detector that minimizes the average error probability is given by the so-called integer least-squares problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{Z}^n} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2, \tag{1.2}$$

where $\|\cdot\|$ denotes the Euclidean norm. Eq. (1.2) is known as the closest vector problem (CVP) in the lattice theory. The lattice generated by a generation matrix $\mathbf{H}$ is [11]

$$\Lambda(\mathbf{H}) = \{\mathbf{H}\mathbf{x} : \mathbf{x} \in \mathcal{Z}^n\}. \tag{1.3}$$

The columns of $\mathbf{H}$ are called basis vectors of $\mathbf{\Lambda}$, and the number n is said to be the dimension of $\mathbf{\Lambda}$. Other applications of CVP include vector quantization and cryptography [11].

Note that (1.1) is a model for real systems. Often, complex signal constellations such as quadrature amplitude modulation (QAM) are used. If (1.1) is considered to be complex, the resulting complex detection problem can be transformed into an equivalent real problem as

$$\tilde{\mathbf{x}} = \arg \min_{\tilde{\mathbf{x}} \in \mathcal{Q}^n} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 \tag{1.4}$$

where

$$\tilde{\mathbf{r}} = \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix}, \tilde{\mathbf{x}} = \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} \tag{1.5}$$

and

$$\tilde{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix}. \tag{1.6}$$

$\mathcal{Q}$ denotes the real constellation after the transformation. For example, a square QAM constellation is transformed to a pulse amplitude modulation (PAM) constellation in (1.4). For a MIMO system with N_T transmitted antennas, we have $n = 2N_T$ in (1.4). Since (1.4) is similar to (1.2), we only discuss how to solve (1.2) in the following using two classic detection algorithms in communications theory. The vertical Bell Laboratories Layered Space-Time Architecture (V-BLAST) detection algorithm [1], which is suboptimal, and the optimal sphere decoding algorithm [2].

1.2.1 V-BLAST algorithm

The V-BLAST detection algorithm [1] consists of nulling and interference cancellation. Nulling is performed by linearly weighting the received symbols to satisfy the zero forcing (ZF) or minimum mean squared error (MMSE) performance criterion. Denoting the i-th column of $\mathbf{H}$ as $\mathbf{h}_i$, the zero-forcing nulling vector $\mathbf{w}_i$, $i = 1, \ldots, n$ is chosen such that

$$\mathbf{w}_i^T \mathbf{h}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{1.7}$$

For interference cancellation, the effect of already-detected symbols can be subtracted from the symbols yet to be detected. This improves the overall performance when the order of detection is chosen carefully. For example, denoting the received vector $\mathbf{r}$ by $\mathbf{r}_1$, if the nulling vector is $\mathbf{w}_1$, the first symbol is then detected by

$$\hat{x}_1 = \arg\min_{x \in \mathcal{Q}} \left| x - \mathbf{w}_1^H \mathbf{r}_1 \right|^2. \tag{1.8}$$

The interference due to $\hat{x}_1$ on the other symbols can be subtracted by taking $\mathbf{r}_2 = \mathbf{r}_1 - \hat{x}_1 \mathbf{h}_1$. Assuming $\hat{x}_1 = x_1$ (i.e., the decision is correct), the next symbol x_2 is then detected using $\mathbf{w}_2$. The detection process consists of n iterations. In the k-th iteration, the signal with maximum post-detection SNR among the remaining $n - k + 1$ symbols is detected, which is known to be the optimal detection order. The post-detection SNR for the k-th detected symbol is given by

$$\rho_k = \frac{E\{|x_k|^2\}}{\sigma_n^2 ||\mathbf{w}_k||^2}. \tag{1.9}$$

From (1.9), maximizing ρ_k is equivalent to minimizing $||\mathbf{w}_k||^2$.

The whole algorithm is described as follows

- Initialization:

$$\mathbf{r}_1 = \mathbf{r} \tag{1.10}$$

$$\mathbf{G}_1 = \mathbf{H}^\dagger \tag{1.11}$$

$$k_1 = \arg\min_j \|(\mathbf{G}_1)_j\|^2 \tag{1.12}$$

- Recursion: for $i = 1$ to n

$$\mathbf{w}_{k_i} = (\mathbf{G}_i)_{k_i} \tag{1.13}$$

$$\hat{x}_{k_i} = \arg\min_{x \in \mathcal{Q}} \left| x - \mathbf{w}_{k_i}^H \mathbf{r}_i \right|^2 \tag{1.14}$$

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \hat{x}_{k_i} (\mathbf{H})_{k_i} \tag{1.15}$$

$$\mathbf{G}_{i+1} = \mathbf{H}_{\bar{k}_i}^{\dagger} \tag{1.16}$$

$$k_{i+1} = \arg\min_{j \notin \{k_1, \ldots, k_i\}} \|(\mathbf{G}_{i+1})_j\|^2 \tag{1.17}$$

where $(\mathbf{A})_i$ is the i-th column of matrix $\mathbf{A}$ and $\mathbf{H}_{\bar{k}_i}$ is obtained by zeroing the $k_1, \ldots, k_i$-th columns of $\mathbf{H}$.

As suggested in [12], given an optimum order $k_1, \ldots, k_n$, V-BLAST detection is equivalent to the zero forcing decision feedback detector (ZF-DFD). Assuming $\mathbf{\Pi}$ is the column permutation matrix obtained from the optimum order, we apply $\mathbf{\Pi}$ to $\mathbf{H}$. As in [12], the filtering matrices in constrained linear detectors and the corresponding constrained ordering can be applied similarly. Let the QR factorization of $\mathbf{G} = \mathbf{H}\mathbf{\Pi}$ be

$$\mathbf{G} = [\mathbf{Q}_1, \mathbf{Q}_2] \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \tag{1.18}$$

where $\mathbf{R}$ is an $n \times n$ upper-triangular matrix, $\mathbf{0}$ is the $(m-n) \times n$ all-zero matrix, $\mathbf{Q}_1$ is an $m \times n$ unitary matrix and $\mathbf{Q}_2$ is an $m \times (m-n)$ unitary matrix. Eq. (1.1) is equivalent to

$$\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{v} \tag{1.19}$$

where $\mathbf{y} = \mathbf{Q}_1^T \mathbf{r}$ and $\mathbf{v} = \mathbf{Q}_1^T \mathbf{n}$ is also an i.i.d. complex Gaussian vector with mean zero and variance σ_n^2. The second description of V-BLAST algorithm is given by

- for $i = n$ to 1

$$\hat{x}_i = \arg\min_{x \in \mathcal{Q}} |y_i - r_{i,i} x|^2 \tag{1.20}$$

$$\mathbf{y} = \mathbf{y} - (\mathbf{R})_i \hat{x}_i \tag{1.21}$$

end

where $r_{i,i}$ is the (i,i)-th entry of $\mathbf{R}$.

1.2.2 Sphere decoding algorithm

Another remarkable algorithm for (1.2) stems from [13], which is well-known as sphere decoding in communication theory [2]. The original SD only tests the lattice points lying inside a hypersphere. Schnorr and Euchner (SE) [14] then suggested an important improvement of the SD by examining the nearest points to the center of the hypersphere first. In wireless communications, the SD first appeared in [2] on lattice code decoding. Since then, the SD has gained popularity in CDMA [15], space-time coding [16] and MIMO systems over frequency selective channels [17]. Its popularity stems from the fact that the SD offers ML decoding at lower complexity, as opposed to the exponential complexity incurred by the exhaustive search. In [3], it has been proven that for a wide range of SNR and lattice dimension, the expected complexity of the SD is polynomial, often cubic in the lattice dimension.

In the following, we use the formulation (1.19). The lattice point $\mathbf{Rx}$ lies in a sphere of radius d if, and only if

$$\|\mathbf{y} - \mathbf{Rx}\|^2 \leq d^2. \tag{1.22}$$

Eq. (1.22) can be written as

$$\sum_{i=1}^{n} \left| y_i - \sum_{j=i}^{n} r_{i,j} x_j \right|^2 \leq d^2 \tag{1.23}$$

where $r_{i,j}$ denotes the (i,j)-th entry of $\mathbf{R}$. The left-hand side (LHS) of the above inequality can be expanded as

$$(y_n - r_{n,n}x_n)^2 + (y_{n-1} - r_{n-1,n}x_n - r_{n-1,n-1}x_{n-1})^2 + \ldots + \sum_{i=1}^{n} \left(y_1 - \sum_{j=1}^{n} r_{1,j} x_j \right)^2 \leq d^2 \tag{1.24}$$

where the first term depends only on x_n, the second term on x_n, x_{n-1} and so forth. Therefore a necessary condition for $\mathbf{Rx}$ to lie inside the sphere is that $d^2 \geq (y_n - r_{n,n}x_n)^2$, which is equivalent to x_n belonging to the interval

$$\left\lceil \frac{-d + y_n}{r_{n,n}} \right\rceil \leq x_n \leq \left\lfloor \frac{d + y_n}{r_{n,n}} \right\rfloor \tag{1.25}$$

where $\lceil \cdot \rceil$ denotes the smallest integer greater than or equal to its argument and $\lfloor \cdot \rfloor$ denotes the largest integer less than or equal to its argument.

For each candidate x_n satisfying the above bound, we define $d_{n-1}^2 = d^2 - (y_n - r_{n,n}x_n)^2$. We can get the following stronger necessary condition for x_{n-1}

$$(y_{n-1} - r_{n-1,n}x_n - r_{n-1,n-1}x_{n-1})^2 \leq d_{n-1}^2 \tag{1.26}$$

which leads to the following bound

$$\left\lceil \frac{-d_{n-1} + y_{n-1} - r_{n-1,n}x_n}{r_{n-1,n-1}} \right\rceil \leq x_{n-1} \leq \left\lfloor \frac{d_{n-1} + y_{n-1} - r_{n-1,n}x_n}{r_{n-1,n-1}} \right\rfloor. \tag{1.27}$$

The SD chooses a candidate for x_{n-1} from the above region. We continue in the same fashion for x_{n-2} and so on. The bounds for x_k are

$$\left\lceil \frac{-d_k + y_k - \sum_{j=k+1}^{n} r_{k,j}x_j}{r_{k,k}} \right\rceil \leq x_k \leq \left\lfloor \frac{d_k + y_k - \sum_{j=k+1}^{n} r_{k,j}x_j}{r_{k,k}} \right\rfloor. \tag{1.28}$$

where $d_k^2 = d_{k+1}^2 - \left(y_{k+1} - \sum_{j=k+1}^{n} r_{k+1,j}x_j\right)^2$. If there is no lattice point within the bounds for x_k, the SD comes back to x_{k+1} and chooses another candidate value from the corresponding region for x_{k+1}. If the SD reaches x_1, the SD finds a candidate lattice point $\mathbf{x}'$ within the hypersphere of radius d. SD checks the value of $\|\mathbf{y} - \mathbf{R}\mathbf{x}'\|^2$. If this value is less than d, we update the radius d which means the search space is limited by the new radius. The above process continues until no further lattice points is found within the hypersphere. The lattice point that achieves the smallest value of $\|\mathbf{y} - \mathbf{R}\mathbf{x}\|^2$ within the hypersphere is deemed as the ML solution. If no point in the sphere is found, the sphere is empty and the search fails. In this case, the initial search radius d must be increased and the search is restarted with the new squared radius. In [3], the authors analyzed the complexity based on the statistical property of the problem. They choose an initial radius such that they can find a lattice point with probability $1 - \epsilon$, and if no point is found, they increase it to a probability of $1 - \epsilon^2$ and so on. The complexity of the SD with such choice of initial radius is [3]

$$C(m, \sigma_n^2, \epsilon) = \sum_{i=1}^{\infty} (1-\epsilon)\epsilon^{i-1} \sum_{k=1}^{m} f_p(k) \sum_{l=0}^{\infty} \gamma\left(\frac{\alpha_i m \sigma_n^2}{2(\sigma_n^2 + l)}, \frac{n-m+k}{2}\right) r_k(l) \tag{1.29}$$

where α_i is chosen such that

$$\gamma\left(\frac{\alpha_i n}{2}, \frac{n}{2}\right) = 1 - \epsilon^i, \quad i = 1, 2, \ldots \tag{1.30}$$

where $f_p(k)$ is the number of elementary operations (additions, subtractions and multiplications) that the SD performs per each visited point in dimension k, $\gamma(\alpha, x)$ is the incomplete gamma function given by

$$\gamma(\alpha, x) = \int_x^{+\infty} t^{\alpha-1} e^{-t} dt, \tag{1.31}$$

and $r_k(l)$ is given by the coefficient of x^n in the expansion

$$\left(1 + 2\sum_{i=1}^{\infty} x^{i^2}\right)^k = 1 + \sum_{l=1}^{\infty} r_k(l) x^l. \tag{1.32}$$

The complete SD algorithm is shown in **Algorithm 1**.

1.3 Thesis Contributions

Although the average time complexity of SD is low at high SNR, the complexity can become large when the SNR is low, or when the problem dimension is high. In addition, for VLSI circuit implementation of the SD, the throughput is limited by the worst-case complexity, which is exponential in the number of variables. Therefore, it is important to develop suboptimal algorithms for large system applications. In Chapter 2, we develop a unified framework for detecting MIMO systems, which includes the well-known algorithms such as ZF-BLAST [1], SD [2], combined ML and ZF-DFD [18] and the B-Chase detector [19] as special cases. We generalize the feedback decoder of Heller [20] for convolutional codes to a new generalized feedback detector (GFD) with three characteristic parameters: window size, step size and branch factor. With different values for these parameters, the GFD achieves different diversity orders between 1 and N and different SNR gains. These parameters also provide a flexible performance-complexity tradeoff. We also consider the relaxation approach to the MIMO detection problem. A class of constrained linear detectors and a class of constrained decision feedback detectors are developed. Moreover, a polynomial constrained detector is proposed and solved using penalty function and differential equations. For high order constellations, we derive a multistage sphere decoding (MSD) algorithm, which exploits the fact that many higher-order signal constellations can naturally be decomposed into several lower-order constellations.

input : $\mathbf{y}$, $\mathbf{R}$, d.
output: The optimal $\mathbf{x}$.

1 Set $k = n$, $d_n^2 = d^2$, $\rho_n = y_n$, $C_m = \infty$;
2 Set $UB_k = \left\lfloor \frac{d_k + \rho_k}{r_{k,k}} \right\rfloor$ and $LB_k = \left\lceil \frac{-d_k + \rho_k}{r_{k,k}} \right\rceil$, $x_k = LB_k - 1$ (Bounds for x_k);
3 $x_k = x_k + 1$(Increase x_k);
4 **if** $x_k \leq UB(x_k)$ **then**
5 | **goto** 10;
6
7 **else**
8 | **goto** 17;
9 **end**
10 $k = k + 1$ (Increase k);
11 **if** $k = n + 1$ **then**
12 | **goto** 31;
13
14 **else**
15 | **goto** 3;
16 **end**
17 **if** $k = 1$ **then**
18 | **goto** 26;
19
20 **else**
21 | $k = k - 1$;
22 | $\rho_k = y_k - \sum_{j=k+1}^{n} r_{k,j} x_j$;
23 | $d_k^2 = d_{k+1}^2 - \left(y_{k+1} - \sum_{j=k+1}^{n} r_{k+1,j} x_j\right)^2$;
24 | **goto** 2;
25 **end**
26 $C = d_n^2 - d_1^2 + (y_1 - r_{1,1} x_1)^2$;
27 **if** $C < C_m$ **then**
28 | $C_m = C$, $\mathbf{x}_m = \mathbf{x}$;
29 **end**
30 **goto** 3;
31 **return** x_m;

Algorithm 1: Sphere Decoding Algorithm

Chapters 3 and 5 focus on the application of efficient detection algorithms to different systems. In Chapter 3, we derive a general decision rule for the ML blind OSTBC decoding in a quasi static (QS) fading channel. Using the linear dispersion property of OSTBC, the decision rule is shown to be a quadratic minimization problem, which can be solved using V-BLAST, SD or the algorithms in Chapter 2. To improve the bandwidth efficiency, a novel approach for totally blind decoding without any pilots is presented using two different constellations. A superimposed training scheme is also presented. Moreover, we give an MMSE channel estimator and derive the Cramér-Rao bound (CRB). Similarly, we develop new blind and semi-blind data detectors for OFDM systems in Chapter 5, which also result in a quadratic form in data symbols. Since the semi-blind detector requires both channel correlation and noise variance, we propose a cyclic-prefix based channel correlation and noise variance estimation algorithm. An enhanced data detector is also derived by noting that for a given least squares (LS) channel estimate, the true channel impulse response (CIR) can be modeled as complex Gaussian with mean being the LS channel estimate itself. The LS channel estimate thus gives a prior on the true channel and averaging the likelihood function over the prior distribution gives the enhanced detector that mitigates the effect of channel estimation errors.

In Chapter 4, we develop a new, efficient detector for DUSTM. Different from the n-dimensional NP-hard problem (1.2), the DUSTM detection problem can be formulated as a one-dimensional NP-hard problem. A fast exact ML detector, called bound-intersection detector (BID), is derived for single symbol detection with diagonal constellations. A key novel feature of the BID is the use of the extended Euclidean algorithm [21], well-known for determining the greatest common divisor (gcd) of two integers, to generate the candidate sets. While the ML-search complexity is exponential in the number of transmit antennas and the data rate, our algorithm, particularly in high SNR, achieves significant computational savings over the naive ML algorithm. We have also developed four BID variants for multiple symbol detection of MSD. The first two are ML and use branch-and-bound (BnB), the third one is suboptimal, which first uses BID to generate a candidate subset and then exhaustively searches over the

reduced space, and the last one generalizes decision-feedback differential detection.

Chapter 6 concludes the thesis and outlines future work in this area.

Chapter 2

Efficient Data Detection for Uncoded MIMO Systems

This chapter presents several efficient data detection algorithms for spatial multiplexing multiple-input multiple-output (MIMO) systems. Section 2.1 introduces the background of the problem. Section 2.2 develops a general framework for tree search based detection algorithms. In Section 2.3, a class of constrained linear detectors and a class of constrained decision feedback detectors are developed, and a polynomial constrained detector is also proposed. A multistage sphere decoder for high order constellations is given in Section 2.4.

2.1 Introduction

2.1.1 Background

As stated in Chapter 1, the use of MIMO systems in a rich scattering wireless channel is capable of providing enormous theoretical capacity improvements without increasing the bandwidth or transmitted power. Because of the promise of extremely high spectral efficiencies, MIMO techniques have attracted considerable interest in the wireless research community and are under consideration for future high-speed wire-

less applications including wireless local area network (LAN) and wireless cellular systems. The Bell-Labs layered space-time (BLAST) architecture is such a MIMO system.

In uncoded MIMO systems, the complexity of the maximum-likelihood detector (MLD) increases exponentially with the number of transmit antennas, making the MLD infeasible. Several reduced-complexity suboptimal detectors have thus been proposed in the literature. The zero-forcing (ZF) decision feedback detector (DFD) with optimal ordering or the V-BLAST detector is proposed in [1] using nulling and interference cancellation. Using nulling based on the minimum mean square error (MMSE) principle, ZF-DFD is extended to MMSE-DFD [22], and this detector offers a compromise between interference suppression and noise enhancement. Since the performance of such detectors is significantly inferior to that of the MLD, and since the complexity of the MLD is substantial, recent research has focussed on developing reduced-complexity high-performance MIMO detectors. For example, in [18], a combined detector (ML-DFD) is proposed to detect the first few symbols using a MLD and the remaining symbols using a ZF-DFD, which prevents the error propagation resulting in a higher diversity order. In [2], the sphere decoder (SD) is proposed as an optimal detection method, which has low complexity in high SNR. However, in low SNR or for systems with a large number of transmit antennas, the SD complexity can be high. The Chase decoder for linear block codes has been adopted for MIMO detection in [19]. The Chase MIMO detector generates a list for the first detected symbol. For each element from the list, a subdetector is applied to the remaining symbols. The vector with the minimum mean square error is chosen as the output. Depending on the type of subdetector, the performance of the Chase detector varies between those of ML and ZF-BLAST. Different SNR gains can be achieved with different list sizes. But the Chase detector achieves a diversity order of 1 or N in an $N \times N$ system, but nothing in between.

In the relaxation approach, the discrete set is embedded in a larger multidimensional continuous space, and the minimization is performed over this continuous space. For example, linear detectors using relaxation approach have been proposed for other

systems. In code-division multiple-access (CDMA) systems, a generalized MMSE (GMMSE) detector is proposed [23], where the binary phase shift keying (BPSK) vectors are relaxed so that they lie inside the smallest hypersphere that contains the unit hypercube. In [24], a tighter relaxation is used to develop a constrained least squares (CLS) detector for orthogonal frequency division multiplexing (OFDM) / spatial division multiple access (SDMA) systems employing unitary signal constellations. In [25], semidefinite relaxation (SDR) has been applied to CDMA systems with BPSK. SDR has been extended to general M-PSK constellations in [26]. However, all these relaxations are loose, which motivates the search for tighter and universal relaxations applicable for any constellations.

In this chapter, we first develop a unified framework for detecting spatial multiplexing systems such as V-BLAST. We reformulate the MIMO detection problem as a set of overlapping subdetection problems; the feedback decoder of Heller [20] for convolutional codes is then extended to the new generalized feedback MIMO detector (GFD) with three characteristic parameters: window size, step size and branch factor. With different values for these parameters, the GFD provides a performance-complexity tradeoff and also yields many well-known algorithms such as the ZF-BLAST [1], the SD [2], the combined ML and ZF-DFD [18] and the B-Chase detector [19] as special cases. Moreover, all such detectors can be explained as tree search algorithms. If linear detectors are used as the subdetector, the GFD also generalizes the L-Chase detector in [19]. A reduced-complexity shared computation technique is also proposed, ensuring the GFD complexity varies between those of the ZF-DFD and the MLD. The symbol error rate of a detector depends on the diversity order and SNR gain. These are derived analytically as a function of the three parameters; for example, the diversity order of the GFD varies between 1 and N. We use the union bound (UB) approach for the symbol error probability for this purpose.

We then extend the idea of constrained detection into uncoded MIMO systems. A class of constrained linear detectors and a class of constrained DFDs are developed. Real constrained linear and decision feedback detectors are proposed for real constellations by suppressing the imaginary interference component. We generalize the

CLS detector of Thoen *et al.* [24] by dividing the signal vector to several subgroups and applying the constant modulus constraints to these subgroups. Similarly, the GMMSE detector is extended to non-constant modulus constellations. New ordering scheme is also proposed using the constrained linear detectors, which maximizes the signal-to-interference and noise ratio (SINR). In V-BLAST, the first detected symbol limits the overall performance. The constrained linear detector and the DFD are thus combined to improve the quality of the first detected symbol and to mitigate the error propagation inherent in the DFD. A polynomial relaxation is also proposed, which can be applied for any constellations. The ML MIMO detection problem is hence reformulated as an equality constrained minimization problem. The constrained optimization problem is solved using a penalty function with Newton method. Because the Newton method may be trapped by local minima, a differential equations algorithm using classical mechanics is proposed to improve the detection performance.

Finally, we develop a multistage sphere decoder for systems using high order constellations. This new multistage sphere decoder exploits the fact that many higher-order signal constellations can naturally be decomposed into several lower-order constellations. We develop a 2-stage SD for a 16-ary quadrature amplitude modulation (16QAM) MIMO system by decomposing 16QAM into two 4QAM constellations. The first stage generates a list of 4QAM vectors. For each of these, the second stage computes an optimal 4QAM vector.

2.1.2 System model

We consider a spatial multiplexing MIMO system with N_T transmit antennas and N_R receive antennas. At the transmitter, source data are mapped into complex symbols from a finite constellation $\mathcal{Q}$. The resulting data stream is partitioned into N_T parallel substreams, each of which is sent though a different transmit antenna over a rich scattering memoryless (flat fading) channel. Each receive antenna receives the signals from all the N_T transmit antennas. The discrete-time baseband received

signals at any time t can thus be written as

$$\mathbf{r} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.1}$$

where $\mathbf{x} = [x_1, \ldots, x_{N_T}]^T$, $x_i \in \mathcal{Q}$ is the transmitted signal vector, $\mathbf{r} = [r_1, \ldots, r_{N_R}]^T$, $r_i \in \mathcal{C}$, is the received signal vector, $\mathbf{H} = [h_{i,j}] \in \mathcal{C}^{N_R \times N_T}$ is the channel matrix, and $\mathbf{n} = [n_1, \ldots, n_{N_R}]^T$, $n_i \in \mathcal{CN}(0, \sigma_n^2)$, is an additive white Gaussian noise (AWGN) vector. The components of $\mathbf{n}$ are identically independent distributed (i.i.d.) complex Gaussian, so are the components of $\mathbf{H}$, i.e., $h_{i,j} \sim \mathcal{CN}(0, 1)$. Note that the linear model (2.1) can also be applied to certain spatially coded MIMO systems, single antenna systems over time and frequency selective channels, intersymbol interference (ISI) channels, and multiuser systems. Consequently, the GFD can also be used for such applications as multiuser detection for CDMA. For brevity, we restrict our considerations to MIMO in this chapter.

For an AWGN channel and i.i.d. source data, and assuming that the channel is perfectly known to the receiver, the MLD that minimizes the average error probability is given by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{Q}^{N_T}} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2. \tag{2.2}$$

Due to the discrete alphabet $\mathcal{Q}$, linear detectors such as least-squares detectors generally do not give the optimal solution. Eq. (2.2) resembles (1.2) except that $\mathcal{Z}$ is replaced by $\mathcal{Q}$. Throughout this section, we assume that the channel is perfectly known to the receiver and $N_T \leq N_R$. If $N_T > N_R$, for constant modulus constellations with unity energy, we have $\mathbf{x}^H\mathbf{x} = N_T$. Since $\sigma_n^2\mathbf{x}^H\mathbf{x}$ is a constant, adding it to the right hand side of (2.2) does not change the optimization problem and we get the equivalent problem as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{Q}^{N_T}} \|\mathbf{r}' - \mathbf{H}'\mathbf{x}\|^2. \tag{2.3}$$

where

$$(\mathbf{H}')^H\mathbf{H}' = \mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}_{N_T}, \quad \mathbf{r}' = (\mathbf{H}')^{-H}\mathbf{H}^H\mathbf{r}. \tag{2.4}$$

For non-constant modulus constellations, when N_T is large, $\mathbf{x}^H\mathbf{x}$ can also be approximated as a constant using the law of large numbers. Therefore (2.3) is still valid in

this case. If N_T is not very large, systems employing M-QAM constellations can also be reformulated as below.

Let a M-QAM constellation and all M-QAM $N \times 1$ vectors be $\mathcal{Q}_M$ and $\mathcal{Q}_M^{N_T}$. We note that any M-QAM ($M = 2^n$) constellation can be represented as a weighted sum of $n/2$ quadrature phase shift keying (QPSK) constellations when n is an even number [27]. That is, for $s \in M$-QAM and $s_i \in$ QPSK, $0 \leq i < n/2$, we have

$$s = \sum_{i=0}^{\frac{n}{2}-1} 2^i \left(\frac{\sqrt{2}}{2}\right) s_i. \tag{2.5}$$

For brevity, we only show how to decompose 16QAM so that our algorithm is applicable. Other M-QAM cases can be derived similarly. Using (2.5), the 16QAM transmit vector $\mathbf{x}$ can be expressed as

$$\mathbf{x} = \sqrt{2}\mathbf{x}_1 + \frac{\sqrt{2}}{2}\mathbf{x}_2 \tag{2.6}$$

where $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q}_4^{N_T}$. Eq. (2.1) can then be represented as

$$\begin{aligned}\mathbf{r} &= \begin{bmatrix} \sqrt{2}\mathbf{H} & \frac{\sqrt{2}}{2}\mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \mathbf{n} \\ &= \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \mathbf{n}. \end{aligned} \tag{2.7}$$

Since $\tilde{\mathbf{x}}^H\tilde{\mathbf{x}}$ is constant, (2.3) can be formulated similarly. It can be readily verified that $\mathbf{H}'$ in (2.3) is an $N_T \times N_T$ full rank matrix. Therefore, our proposed GFD and the former detectors [1,18,19,28] can also be applied to the equivalent problem (2.4).

Before proceeding, we first transform MLD (2.2). Column reordering can be applied to $\mathbf{H}$ by using V-BLAST or other ordering schemes [22,28] and the resulting matrix is $\mathbf{G} = \mathbf{H}\mathbf{\Pi}$, where $\mathbf{\Pi}$ is the column permutation matrix. Following the same QR factorization as in (1.18), we find (2.2) is equivalent to

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in Q^{N_T}} \|\mathbf{y} - \mathbf{R}\mathbf{x}\|^2 \tag{2.8}$$

where $\mathbf{y}$ and $\mathbf{R}$ are defined in (1.19) and $\mathbf{R}$ is an upper triangular matrix. Eq. (2.8) is the basis for our GFD and the former detectors [1,18,19,28].

2.2 Unified Framework for Tree Search Based Detection

2.2.1 Feedback decoding

There are three classical decoding algorithms for convolutional codes [29]: Viterbi decoding, sequential decoding and feedback decoding. The Viterbi algorithm achieves optimal maximum-likelihood decoding, but the complexity grows exponentially with the size of the memory of the encoder. Sequential decoding can however perform near ML, but with significantly reduced complexity. Moreover, the SD can be interpreted as a chain of sequential decoders [28]. Feedback decoding due to Heller [20], on the other hand, sacrifices performance in exchange for complexity reduction. However, Heller's feedback decoder has not been applied to MIMO detection to the best of our knowledge. It not only provides a unified framework for describing virtually all existing MIMO detectors, but also enables new ones. We use the term *generalized feedback detector* to differentiate from Heller's feedback decoding for convolutional codes.

We now extend the feedback decoding algorithm to MIMO detection in (2.8). We start by considering only one parameter: *window size* (w). The cost metric to be minimized in (2.8) can be written in scalar form as

$$\sum_{i=1}^{N_T} \left(\left| y_i - \sum_{j=i}^{N_T} r_{i,j} x_j \right|^2 \right) \tag{2.9}$$

where $r_{i,j}$ is the (i,j)-th entry of matrix $\mathbf{R}$. Since $\mathbf{R}$ is upper-triangular, the i-th term in (2.9) only depends on $x_i, \ldots, x_{N_T}$, $1 \leq i \leq N_T$. The feedback detector can be considered as a sliding window algorithm. The detector starts from x_{N_T}. When hard decisions have been made on x_i as $\hat{x}_i$, $k < i \leq N_T$, the detector makes a decision on x_i based on the metric computed from y_{k-w+1} to y_k, where w is a preselected positive integer, and we call it window size. We first determine the subvector $\mathbf{x}^{(k)} =$

$[x_{k-w+1}, \ldots, x_k]^T$ using

$$\hat{\mathbf{x}}^{(k)} = \arg\min_{\mathbf{x}^{(k)} \in Q^w} \sum_{i=k-w+1}^{k} \left(\left| y_i - \sum_{j=i}^{k} r_{i,j} x_j - \sum_{j=k+1}^{N_T} r_{i,j} \hat{x}_j \right|^2 \right). \tag{2.10}$$

In principle, exhaustive search can solve (2.10), and in the resulting $\hat{\mathbf{x}}^{(k)}$, although contain w decisions, we discard $w-1$ of them in order to improve the overall error rate. Consequently, the hard decision on x_k is made to be $\hat{x}_k = \hat{\mathbf{x}}^{(k)}(w)$, where $\hat{\mathbf{x}}^{(k)}(w)$ denotes the w-th element in $\hat{\mathbf{x}}^{(k)}$. The same procedure is performed for x_{k+1} based on the metric computed from y_{k-w} to y_{k+1}. When the window size is unity ($w = 1$), the feedback detector reduces to the ZF-DFD or the V-BLAST detector, and incidentally, the equivalence between these two detectors has been noted in [12]. When the window size increases to N_T, the GFD reduces to the MLD. When window size varies between 1 and N_T, the performance of the GFD is between those of the MLD and the ZF-DFD. Our basic algorithm can be considered as a sliding window detector of size w, making a decision on each x_i based on the minimum metric within the window.

To allow for more performance flexibility, we introduce another parameter *step size* and allow the window size to change (possibly) in each stage. Also starting from x_{N_T}, instead of making single decision in each stage, we detect the group $x_{l_k-s_k+1}, \ldots, x_{l_k}$ at the k-th stage, where $l_k = N_T - \sum_{k'=1}^{k-1} s_{k'}$ and $s_{k'}$ is the step size at the k'-th stage. The subvector $\mathbf{x}^{(k)} = [x_{l_k-w_k+1}, \ldots, x_{l_k}]^T$ is detected using

$$\hat{\mathbf{x}}^{(k)} = \arg\min_{\mathbf{x}^{(k)} \in Q^{w_k}} \sum_{i=l_k-w_k+1}^{l_k} \left(\left| y_i - \sum_{j=i}^{l_k} r_{i,j} x_j - \sum_{j=l_k+1}^{N_T} r_{i,j} \hat{x}_j \right|^2 \right) \tag{2.11}$$

where w_k is the window size in the k-th stage. We choose $[\hat{x}_{l_k-s_k+1}, \ldots, \hat{x}_{l_k}]^T = \hat{\mathbf{x}}^{(k)}(w_k - s_k + 1 : w_k)$. When proceeding to the $(k+1)$-th stage, the sliding window is shifted by s_k, and the window size is changed to w_{k+1}. The group of symbols $x_{l_{k+1}-s_{k+1}+1}, \ldots, x_{l_{k+1}}$ is decided within the new window. If the detector has K stages, we have $\sum_{k=1}^{K} s_k = N_T$. The basic algorithm with different w_k and s_k is illustrated in Fig. 2.1, where the shadowed regions correspond to hard decisions.

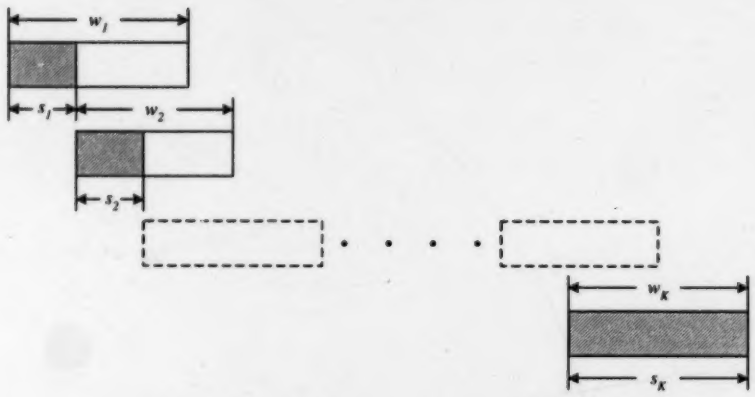

Fig. 2.1. Basic algorithm illustration of the GFD with different w_k and s_k.

In (2.11), hard decisions are made on the components of $\mathbf{x}^{(k)}$. However, if the group size s_k is large, the error performance degrades rapidly. To overcome this, we generate a list of candidates for $\mathbf{x}^{(k)}$ that minimize (2.11), instead of returning the best candidate. The size of the list is b_k named *branch factor*. The corresponding different prefix vectors $\mathbf{x}^{(k)}(w_k - s_k + 1 : w_k)$ are stored in another list $\mathcal{L}_k$ with q_k elements. Since some $\mathbf{x}^{(k)}$ may have the same prefix, we have $1 \leq q_k \leq b_k$. Similarly, b_k may vary in different stages. We can thus obtain $\mathbf{x}$ by using

$$\hat{\mathbf{x}} = \underset{[x_{N_T-s_1+1},\ldots,x_{N_T}]^T \in \mathcal{L}_1,\ldots,[x_1,\ldots,x_{s_K}]^T \in \mathcal{L}_K}{\arg\min} \|\mathbf{y} - \mathbf{R}\mathbf{x}\|^2. \tag{2.12}$$

If $b_k = 1$ for $k = 1, \ldots, K$, (2.12) reduces to the detector in the second case. If $b_k = 1$, $s_k = 1$ and $w_k = w$, (2.12) reduces to the original feedback detection algorithm. Note in the K-th stage, we have $b_K = 1$ and $s_K = w_K$ imposed by the end state condition.

Both (2.11) and (2.12) entail an exhaustive search in a reduced space. The SD [28] efficiently solves (2.11) and (2.12). In particular, the Schnorr and Euchner SD (SESD) [14] removes the dependence on the initial radius and hence may be preferred. When $b_k \neq 0$, the list sphere decoder (LSD) [30] can be used to create the candidate list. If not enough candidates have been found, the radius is increased and the LSD searches again until a b_k element list has been formed.

Clearly, if $w_k = 1$, $s_k = 1$ and $b_k = 1$, the GFD reduces to ZF-DFD [1] and if $w_k = N_T$, $s_k = N_T$ and $b_k = 1$, it becomes SD [2]. When $w_1 = p$, $w_k = 1$ $(k > 1)$,

$s_1 = p$, $s_k = 1$ $(k > 1)$ and $b_k = 1$, the GFD reduces to a combination of MLD and ZF-DFD [18]. When $w_k = 1$, $s_k = 1$, $b_1 = q$ and $b_k = 1$ $(k > 1)$, the GFD becomes the B-Chase detector in [19] using a different column permutation matrix. Table 2.1 summarizes the relationship between these detectors and the GFD for different parameter values.

Remarks:

- In the basic algorithm, we propose using the SD to solve the subproblem (2.11) in each window. Of course, the complexity can be reduced by replacing the SD with a suboptimal detector such as LS, MMSE or GMMSE [23], CLS [24], just to name a few. Similarly, the last s_k symbols are detected by making hard decisions. Furthermore, the SD and linear detectors can be used interchangeably. For example, as shown in Section 2.2.4, since the detection of the first few symbols is critical to the overall system performance. The SD can be used to detect the symbols in the first few windows, and then linear detectors can be applied to reduce the overall complexity. Specifically, if $w_1 = 1$, $w_2 = N_T - 1$, $s_1 = 1$, $s_2 = N_T - 1$, $b_1 = q$, $b_2 = 1$ and exhaustive search is used in the first window, linear detector is used in the second window, our GFD reduces to the L-Chase detector in [19]. Allowing linear detectors, our GFD thus generalizes the L-Chase detector and provides more flexible complexity and performance tradeoff.

- Our proposed GFD can be extended to decoding convolutional codes. Moreover, it can also be readily modified to perform joint detection and decoding in coded MIMO systems employing both vertical and horizontal coding architectures [31].

2.2.2 New tree interpretation of MIMO detection

All MIMO detection algorithms can be interpreted as performing a search through a tree, which is unavailable in the literature to the best of our knowledge. The detectors traverse through a $|\mathcal{Q}|$-ary tree of N_T levels, where $|\mathcal{Q}|$ is the cardinality

TABLE 2.1

Relationship to the generalized feedback detector.

Detector	Window size w_k	Step size s_k	Branch factor b_k
ZF-DFD	$w_k = 1$	$s_k = 1$	$b_k = 1$
SD	$w_k = N_T$	$s_k = N_T$	$b_k = 1$
ML-DFD	$w_1 = p$, $w_k = 1, k > 1$	$s_1 = p$, $s_k = 1, k > 1$	$b_k = 1$
B-Chase	$w_k = 1$	$s_k = 1$	$b_1 = q$, $b_k = 1, k > 1$

of constellation $\mathcal{Q}$. Except for the leaf nodes, there are $|\mathcal{Q}|$ branches stemming from each node, and each branch is labelled by an element from $\mathcal{Q}$. A node at the k-th level is assigned a metric

$$m_k(\mathbf{x}_k) = \left| y_{N_T+1-k} - \sum_{j=N_T+1-k}^{N_T} r_{N_T+1-k,j} x_j \right|^2 \tag{2.13}$$

where $\mathbf{x}_k = [x_{N_T+1-k}, \ldots, x_{N_T}]^T$ are the symbols labelling the path from the root to this node. The accumulated path metric associated with path $\mathbf{x}_k$ is thus defined as

$$c(\mathbf{x}_k) = \sum_{i=N_T+1-k}^{N_T} m_i(\mathbf{x}_i) = \sum_{i=N_T+1-k}^{N_T} \left| y_i - \sum_{j=i}^{N_T} r_{i,j} x_j \right|^2 . \tag{2.14}$$

The MLD performs an exhaustive tree search by computing the accumulated path metric (2.14) for all possible tree paths from the root to leaf nodes. At the leaf node of the tree, the path with minimum accumulated metric is selected as the ML solution.

Instead of exhaustive search by visiting all of the nodes, the SD only explores a subset of the tree. Using a global cost upper bound, the SD explores only those nodes whose accumulated path metric is less than the global bound and discards other nodes. When the SD reaches a leaf node, if its cost is less than the global upper bound, the latter is replaced by the former. The order of the processing of the child nodes of a node can impact the complexity. In the SESD, the processing order depends on the accumulated metric. Thus, the child node with minimum accumulated metric is

expanded by the SESD first. This ordering ensures that whenever the SESD reaches a leaf node, its path cost is as small as possible, and thus the global bound is reduced rapidly.

In contrast to the SD, DFDs are greedy tree search algorithms. Since the DFD only expands the child node with minimum accumulated path metric. Therefore, only one path is traversed through the search tree, which has complexity linear in N_T but also a large performance loss.

Our proposed GFD only expands a set of partial trees instead of the full tree in the MLD. At the k-th stage, $x_{l_k}, \ldots, x_{N_T}$ are assigned values. The GFD searches through a w_k level tree stemming from x_{l_k-1}. Then b_k best partial paths from the root to the leaf nodes of the partial tree with minimum accumulated metric are chosen. $x_{l_k-1}, \ldots, x_{l_k-s_k}$ symbols are chosen corresponding to each partial path. The window shifts s_k symbols and a partial tree is searched again. In fact, our GFD forms a reduced tree with K levels. At the k-th level, there are q_k branches from a node, which is assigned a metric

$$\tilde{m}_k(\mathbf{x}_k) = \sum_{i=l_k-s_k}^{l_k-1} \left| y_i - \sum_{j=i}^{N_T} r_{i,j} x_j \right|^2 \tag{2.15}$$

where $\mathbf{x}_k = [x_{l_k-s_k}, \ldots, x_{N_T}]^T$ are the symbols labelling the path from the root to this node. The SD and DFDs can be applied to the reduced tree. If $b_k = 1$, the GFD reduces to ZF-DFD, or the SD can be used to search through the new tree. The branch factor b_k relates to the number of branches in the new tree.

Fig. 2.2 illustrates the GFD on a tree with 4 levels (i.e., $N_T = 4$) and binary phase shift keying (BPSK). The left side is the full tree expanded by exhaustive search. In the GFD, we choose $w_1 = 3$, $s_1 = 2$ and $b_1 = 2$. Exhaustive search or LSD is used to traverse the first partial tree; $[x_4, x_3, x_2]^T = \{[+1, +1, -1]^T, [-1, +1, +1]^T\}$ are the b_1 subvectors that make the partial accumulated metric from x_4 to x_2 a minimum. For each subvector, $[x_4, x_3]^T$ is stored in a list $\mathcal{L}_1 = \{[+1, +1]^T, [-1, +1]^T\}$. The b_1 elements form b_1 branches in the new tree in the right side of Fig. 2.2. In the second stage, the window is shifted by $s_1 = 2$ to the second rectangle, and window size, step size and branch factor are changed to $s_2 = w_2 = 2$ and $b_2 = 1$. Therefore, the original

4-level tree with 16 leaf nodes reduces to a 2-level tree with 2 leaf nodes using GFD. SD or SESD can be used to find the path with minimum accumulated metric in the reduced tree. The corresponding path of the new tree is the output sequence. The ZF-DFD just expands a single branch tree as shown in Fig. 2.2 (c).

The B-Chase detector [19] with list size q and with a ZF-DFD subdetector forms a 2-level tree with q leaf nodes; exhaustive search finds the minimum path metric in the tree.

Our GFD forms different trees with different parameters w_k, s_k and b_k, As will be shown in Section 2.2.4, different parameter set provides different performance and complexity.

2.2.3 Computation sharing technique

The next section shows that the GFD bridges the gap between ZF-DFD and MLD. Another important issue is that whether our GFD also has complexity between ZF-DFD and ML. We first look at a simple case with $b_k = 1$. The discussion can be readily extended to other cases. A detail analysis of the worst complexity is given in Section 2.2.5. The basic GFD constitutes K SDs and a ZF-DFD on a K-level tree. We denote the complexities of SD and DFD for an $n \times n$ system as $C_{\text{SD}}(n)$ and $C_{\text{DFD}}(n)$, respectively. Thus, we can obtain the complexity of the GFD as

$$C_{\text{GFD}}(N_T) = \sum_{k=1}^{K} C_{\text{SD}}(w_k) + C_{\text{DFD}}(K). \tag{2.16}$$

In low SNR, $C_{\text{SD}}(w_k) = O(|Q|^{w_k})$ and $C_{\text{GFD}}(N_T) < C_{\text{SD}}(N_T)$. But, in high SNR, $C_{\text{SD}}(w_k) \simeq w_k^3$ and it is possible that $C_{\text{GFD}}(N_T) > C_{\text{SD}}(N_T)$.

To reduce the complexity gap between the GFD and the SD, we introduce a shared computation technique that eliminates redundant computations in the basic GFD. In the k-th stage, if $w_k \neq s_k$, there will be $w_k - s_k$ symbols' overlap between the k-th window and the $(k+1)$-th window. The basic GFD uses two SDs in the two windows. However, due to the overlap, some nodes of the partial tree in the $(k+1)$-th window have been visited by the sphere decoder in the k-th window. We thus propose to store

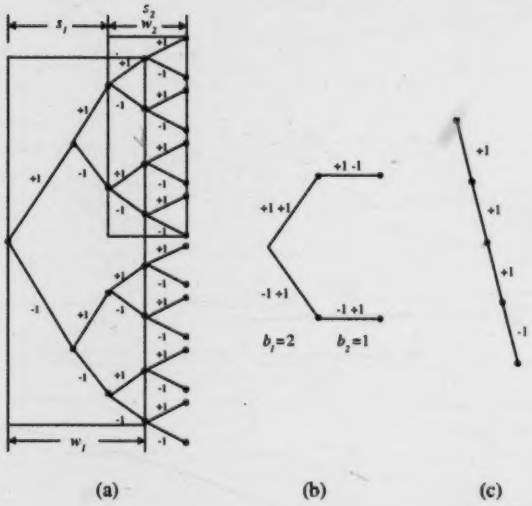

Fig. 2.2. Comparison of exhaustive search, GFD and ZF-DFD in tree representation for $N_T = N_R = 4$ and BPSK.

all the metrics for the visited nodes in the k-th SD to avoid repeated computation. The *tree* is a suitable data structure for visited nodes storage since the SD forms a tree during the search. When using the SD for the $(k+1)$-th partial tree, the SD traverses the search tree and the storage tree at the same time. Only the subtree corresponding to the selected $[x_{l_k-s_k+1}, \ldots, x_{l_k}]^T$ is kept and the other subtrees are pruned. If the node has not been visited as indicated in the storage tree, the branch metric is computed, and the node is added into the storage tree.

When $b_k > 1$, LSD must be used. Given an initial radius, if less than b_k candidates have been found within the radius, LSD needs to enlarge the radius and search again until b_k points are found. Similarly, enlarging the radius and repeating the search revisits several nodes that have already been visited. The storage tree can also be created for LSD to reduce repeated computation. However, shared computation needs more memory, which is exponential in w_k in the worst case. It provides a memory-complexity tradeoff.

2.2.4 Performance analysis

In this section, we consider the case of $N_T = N_R$. In the GFD, we first assume $w_k = w$, $s_k = s$ and $b_k = 1$. All the analysis and results can be easily extended to the case $N_T \neq N_R$ and general GFDs. From [32], the squared norm of the entries of upper-triangular matrix $\mathbf{R}$ have χ^2 distribution with different degrees of freedom, specifically, $|r_{i,i}|^2 \sim \chi^2(2i)$, for $i = 1, \ldots, N_T$ and $|r_{i,j}|^2 \sim \chi^2(2)$, for $j > i$, where $\chi^2(k)$ denotes the chi-squared distribution with k degrees of freedom.

In the first stage of the GFD, we perform ML decoding with $\mathbf{y}_1 = \mathbf{R}_1\mathbf{x}_1 + \mathbf{v}_1$ to detect $\mathbf{x}_1 = [x_{N_T-w+1}, \ldots, x_{N_T}]^T$, where $\mathbf{y}_1 = [y_{N_T-w+1}, \ldots, y_{N_T}]^T$, $\mathbf{v}_1 = [v_{N_T-w+1}, \ldots, v_{N_T}]^T$ and $\mathbf{R}_1 = \mathbf{R}(N_T - w + 1 : N_T, N_T - w + 1 : N_T)$. Denote the block error event in the j-th window by E_j. The union bound for the block error probability of $\mathbf{x}_1$, $P_b(E_1)$, is given by

$$P_b(E_1) \leq \underset{\mathbf{R}_1}{E} \underset{\mathbf{x}_1^{(1)}}{E} \left[\sum_{\mathbf{x}_1^{(2)} \neq \mathbf{x}_1^{(1)}} P\left(\hat{\mathbf{x}}_1 = \mathbf{x}_1^{(2)} | \mathbf{x}_1^{(1)}, \mathbf{R}_1\right) \right] \tag{2.17}$$

where $\mathbf{x}_1^{(2)}$ are all the possible vectors other than $\mathbf{x}_1^{(1)}$, and

$$P\left(\hat{\mathbf{x}}_1 = \mathbf{x}_1^{(2)} | \mathbf{x}_1^{(1)}, \mathbf{R}_1\right) = Q\left(\sqrt{\left\|\mathbf{R}_1\left(\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right)\right\|^2 / 2\sigma_n^2}\right), \tag{2.18}$$

where $Q(\cdot)$ is the Q-function. Let $\mathbf{H}_2$ be a $w \times w$ matrix with complex Gaussian entries and its QR decomposition be $\mathbf{H}_2 = \mathbf{Q}_2\mathbf{R}_2$. Since $\mathbf{R}_2$ has the same distribution as $\mathbf{R}_1$, $\|\mathbf{R}_2\left(\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right)\|^2$ has the same distribution as $\|\mathbf{R}_1\left(\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right)\|^2$. Considering $\mathbf{Q}_2$ is unitary, we have

$$\left\|\mathbf{R}_2\left(\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right)\right\|^2 = \left\|\mathbf{Q}_2\mathbf{R}_2\left(\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right)\right\|^2 = \left\|\mathbf{H}_2\left(\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right)\right\|^2 = a_1\|\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\|^2 \tag{2.19}$$

where $a_1 \sim \chi^2(2w)$ and the last equation is due to the linear combination of independent Gaussian variables is also Gaussian. Eq (2.18) thus reduces to

$$P\left(\hat{\mathbf{x}}_1 = \mathbf{x}_1^{(2)} | \mathbf{x}_1^{(1)}, a_1\right) = Q\left(\sqrt{a_1\left\|\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\right\|^2 / 2\sigma_n^2}\right). \tag{2.20}$$

Using the Chernoff bound for the Q-function, $P_b(\mathbf{x}_1)$ can be bounded as

$$\begin{aligned} P_b(E_1) &\leq \underset{a_1}{E} \underset{\mathbf{x}_1^{(1)}}{E} \left[\sum_{\mathbf{x}_1^{(2)}} \exp\left(-a_1 \left\| \mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)} \right\|^2 / 4\sigma_n^2 \right) \right] \\ &= \underset{\mathbf{x}_1^{(1)}}{E} \sum_{\mathbf{x}_1^{(2)}} \frac{1}{\left(1 + \frac{\left\| \mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)} \right\|^2}{4\sigma_n^2} \right)^w} \\ &\leq \left(\frac{|\mathcal{Q}|}{1 + d_{\min}^2 / 4\sigma_n^2} \right)^w \end{aligned} \tag{2.21}$$

where $d_{\min}$ is the minimum distance between $\mathbf{x}_1^{(1)}$ and $\mathbf{x}_1^{(2)}$. In high SNR, a block error results from a single symbol error in $\mathbf{x}_1$. Therefore, the bound for the symbol error probability is $P_s(x_i) = P_b(E_1)/w$, $i = N_T - s + 1, \ldots, N_T$ and the bound for the block error probability of $\tilde{\mathbf{x}}_1 = [x_{N_T-s+1}, \ldots, x_{N_T}]^T$ is $P_b(\tilde{E}_1) = P_b(E_1)s/w$, where $\tilde{E}_1$ is the block error event in $\tilde{\mathbf{x}}_1$. We also bound the block error probability for the second window as

$$\begin{aligned} P_b(E_2) =& P(E_2|\tilde{E}_1^c) \cdot P(\tilde{E}_1^c) + P(E_2|\tilde{E}_1) \cdot P_b(\tilde{E}_1) \\ \leq& P(E_2|\tilde{E}_1^c) + P_b(\tilde{E}_1). \end{aligned} \tag{2.22}$$

If $\tilde{\mathbf{x}}_1$ is correct, we cancel it from $\mathbf{y}$ and perform ML detection with $\mathbf{y}_2 = \mathbf{R}_2\mathbf{x}_2 + \mathbf{v}_2$, where $\mathbf{x}_2 = [x_{N_T-s-w+1}, \ldots, x_{N_T-s}]^T$, where $\mathbf{y}_2 = [y_{N_T-s-w+1}, \ldots, y_{N_T-s}]^T$, $\mathbf{v}_2 = [v_{N_T-s-w+1}, \ldots, v_{N_T-s}]^T$ and $\mathbf{R}_2 = \mathbf{R}(N_T-s-w+1 : N_T-s, N_T-s-w+1 : N_T-s)$. Similarly to (2.21), we can obtain

$$P(E_2|\tilde{E}_1^c) \leq \left(\frac{|\mathcal{Q}|}{1 + d_{\min}^2 / 4\sigma_n^2} \right)^{w+s}. \tag{2.23}$$

In general, in the k-th window, we perform ML decoding with $\mathbf{y}_k = \mathbf{R}_k\mathbf{x}_k + \mathbf{v}_k$ to detect $\mathbf{x}_k = [x_{N_T-(k-1)s+1}, \ldots, v_{N_T-(k-1)s+w}]^T$, where $\mathbf{y}_k = [y_{N_T-(k-1)s+1}, \ldots, v_{N_T-(k-1)s+w}]^T$, $\mathbf{v}_k = [v_{N_T-(k-1)s+1}, \ldots, v_{N_T-(k-1)s+w}]^T$ and $\mathbf{R}_k = \mathbf{R}(N_T - (k-1)s + 1 : N_T - (k-1)s + w, N_T - (k-1)s + 1 : N_T - (k-1)s + w)$. The union bound for the event E_k conditional on $\left(\cup_{i=1}^{k-1} \tilde{E}_i \right)^c$ is given by

$$P\left(E_k | \left(\cup_{i=1}^{k-1} \tilde{E}_i \right)^c \right) \leq \underset{\mathbf{x}_k^{(1)}}{E} \left[\sum_{\mathbf{x}_k^{(2)}} P\left(\hat{\mathbf{x}}_k = \mathbf{x}_k^{(2)} | \mathbf{x}_k^{(2)}, \mathbf{R}_k \right) \right] \tag{2.24}$$

where $\mathbf{x}_k^{(2)}$ are all the possible vectors other than $\mathbf{x}_k^{(1)}$, and

$$P\left(\hat{\mathbf{x}}_k = \mathbf{x}_k^{(2)} | \mathbf{x}_k^{(1)}, \mathbf{R}_k\right) = Q\left(\sqrt{\left\|\mathbf{R}_k\left(\mathbf{x}_k^{(2)} - \mathbf{x}_k^{(1)}\right)\right\|^2 / 2\sigma_n^2}\right). \tag{2.25}$$

Let $\mathbf{H}_2$ be a $(k-1)s+w \times (k-1)s+w$ matrix with each entry complex Gaussian and its QR decomposition be $\mathbf{H}_2 = \mathbf{Q}_2\mathbf{R}_2$. Define two new vectors $\tilde{\mathbf{x}}_k^{(1)} = [(\mathbf{x}_k^{(1)})^H, \mathbf{0}_{(k-1)s\times 1}]^H$ and $\tilde{\mathbf{x}}_k^{(2)} = [(\mathbf{x}_k^{(2)})^H, \mathbf{0}_{(k-1)s\times 1}]^H$. Since $\mathbf{R}_2((k-1)s+1:(k-1)s+w, (k-1)s+1:(k-1)s+w)$ has the same distribution as $\mathbf{R}_k$, $\left\|\mathbf{R}_2\left(\tilde{\mathbf{x}}_k^{(2)} - \tilde{\mathbf{x}}_k^{(1)}\right)\right\|^2$ has the same distribution as $\left\|\mathbf{R}_k\left(\mathbf{x}_k^{(2)} - \mathbf{x}_k^{(1)}\right)\right\|^2$. Considering $\mathbf{Q}_2$ is unitary, we have

$$\left\|\mathbf{R}_2\left(\tilde{\mathbf{x}}_k^{(2)} - \tilde{\mathbf{x}}_k^{(1)}\right)\right\|^2 = \left\|\mathbf{Q}_2\mathbf{R}_2\left(\tilde{\mathbf{x}}_k^{(2)} - \tilde{\mathbf{x}}_k^{(1)}\right)\right\|^2 \overset{f}{=} \left\|\mathbf{H}_2\left(\tilde{\mathbf{x}}_k^{(2)} - \tilde{\mathbf{x}}_k^{(1)}\right)\right\|^2 = a_k\|\mathbf{x}_1^{(2)} - \mathbf{x}_1^{(1)}\|^2 \tag{2.26}$$

where $a_k \sim \chi^2(2(k-1)s+2w)$ and the last equation is due to the definition of $\tilde{\mathbf{x}}_k^{(1)}$ and $\tilde{\mathbf{x}}_k^{(2)}$ and linear combination of independent Gaussian variables is also Gaussian. Similar to (2.21), we have

$$P\left(E_k | \left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right) \leq \left(\frac{|\mathcal{Q}|}{1 + d_{\min}^2/4\sigma_n^2}\right)^{w+(k-1)s}. \tag{2.27}$$

Using the total probability theorem [32] and noting that the events $(\cup_{i=1}^{k-1}\tilde{E}_i)^c$, $\tilde{E}_{k-1} \cap (\cup_{i=1}^{k-2}\tilde{E}_i)^c, \ldots, \tilde{E}_1$ are exclusive, we have

$$\begin{aligned}
P_b(E_k) =& P\left(E_k|\tilde{E}_1\right)P\left(\tilde{E}_1\right) + P\left(E_k|\tilde{E}_2 \cap \tilde{E}_1^c\right)P\left(\tilde{E}_2 \cap \tilde{E}_1^c\right) + \cdots \\
&+ P\left(E_k|\tilde{E}_{k-1} \cap \left(\cup_{i=1}^{k-2}\tilde{E}_i\right)^c\right)P\left(\tilde{E}_{k-1} \cap \left(\cup_{i=1}^{k-2}\tilde{E}_i\right)^c\right) \\
&+ P\left(E_k|\left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right)P\left(\left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right) \\
\overset{a}{\leq}& P\left(\tilde{E}_1\right) + P\left(\tilde{E}_2 \cap \tilde{E}_1^c\right) + \cdots + P\left(E_k|\left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right)P\left(\left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right) \\
\overset{b}{=}& \sum_{j=1}^{k-1} P\left(\tilde{E}_j|\cap_{i=1}^{j-1}\tilde{E}_i^c\right)P\left(\cap_{i=1}^{j-1}\tilde{E}_i^c\right) + P\left(E_k|\left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right)P\left(\cap_{i=1}^{k-1}\tilde{E}_i^c\right)
\end{aligned} \tag{2.28}$$

where the inequality (a) is obtained by assuming that given an error event on $\tilde{E}_i$, $1 \leq i < k$, the event E_k has probability 1. This assumption is precise at high SNR. The equality (b) comes from the Bayes' theorem [32]. Using the chain rule for

conditional probability, (b) can be computed in a closed-form. However, this closed-form is complicated and does not show insight into the diversity order and SNR gain of the detection algorithm. Note that in high SNR, $P\left(\cap_{i=1}^{j-1}\tilde{E}_i^c\right)$ is close to 1. We can thus further simplify $P_b(E_k)$ as

$$P_b(E_k) \leq P\left(E_k|\left(\cup_{i=1}^{k-1}\tilde{E}_i\right)^c\right) + \sum_{j=1}^{k-1} P\left(\tilde{E}_j|\cap_{i=1}^{j-1}\tilde{E}_i^c\right). \tag{2.29}$$

Similarly, in high SNR, we have

$$P(\tilde{E}_j|\cap_{i=1}^{j-1}\tilde{E}_i^c) = \frac{s}{w}P_b(E_j|\cap_{i=1}^{j-1}\tilde{E}_i^c) = \frac{s}{w}P_b(E_1)\epsilon^{j-1} \tag{2.30}$$

where $\epsilon = \left(\frac{|\mathcal{Q}|}{1+d_{\min}^2/4\sigma_n^2}\right)^s$. Substituting (2.30) into (2.29), we can obtain

$$\begin{aligned} P_b(E_k) \leq & P_b(E_1)\epsilon^{k-1} + \frac{s}{w}P_b(E_1)\left(1+\epsilon+\cdots+\epsilon^{k-2}\right) \\ \leq & \frac{s}{w}P_b(E_1)(1+\delta). \end{aligned} \tag{2.31}$$

In high SNR, where δ is a small positive number and $(s/w)P_b(E_1)$ dominates the upper bound (2.31). Therefore, the average symbol error probability can be approximated as

$$\begin{aligned} P_s = & \frac{1}{N_T}\left(\frac{s}{w}P_b(E_1) + \frac{s}{w}\sum_{k=2}^{K-1}P_b(E_k) + P_b(E_K)\right) \\ \simeq & \frac{1}{N_T}\left(\frac{2s}{w} + (K-2)\frac{s^2}{w^2}\right)\left(\frac{|\mathcal{Q}|}{1+d_{\min}^2/4\sigma_n^2}\right)^w. \end{aligned} \tag{2.32}$$

Eq. (2.32) shows that the GFD has diversity order w and different s and w provide different SNR gains. The union bound (2.32) is loose and the actual SNR gain may be different. From the performance analysis, it can be verified that if w_k's are different, the performance of GFD is limited by the first window size and the diversity order of GFD is w_1. b_k only determines the SNR gain.

When $b_k > 1$, more SNR gain can be achieved. For the B-Chase detector, the effective SNR gain is defined in [19]. Here we take a different approach by calculating the upper bound on the symbol error probability. We take the QPSK for example and $w = 1$, $\mathcal{Q} = \{e^{-j\pi k/2}|k = 0,1,2,3\}$. We assume that 1 is transmitted without loss of generality. We consider the error event in the first stage E_1. The ML detection

for x_{N_T} reduces to detecting the phase of $y_{N_T} r^*_{N_T,N_T} = ae^{j\theta} = X + jY$. If $b_1 = 1$, x_{N_T} is correctly detected if $|\theta| < \pi/4$. But, when $b_1 = 2$, x_{N_T} is correct if the list contained 1 or equivalently $|\theta| < \pi/2$. Similarly, for $b_1 = 3$, the correct decision region becomes $|\theta| < 3\pi/4$. Given r_{N_T}, $X \sim \mathcal{N}(|r_{N_T}|^2, \sigma_n^2 |r_{N_T}|^2)$ and $Y \sim \mathcal{N}(0, \sigma_n^2 |r_{N_T}|^2)$. The probability density function (pdf) of $\theta = \arctan(Y/X)$ is given by [32]

$$f(\theta) = \frac{1}{2\pi} e^{-\frac{|r_{N_T}|^2}{2\sigma_n^2}} + \frac{|r_{N_T}|}{\sqrt{2\pi}\sigma_n} \cos\theta e^{-\frac{|r_{N_T}|^2}{2\sigma_n^2}\sin^2\theta} Q\left(-\frac{|r_{N_T}|}{\sigma_n}\cos\theta\right). \tag{2.33}$$

The error probability of E_1 can then be obtained as

$$P_b(E_1) = 1 - \int_{-\beta}^{\beta} f(\theta) \mathrm{d}\theta \tag{2.34}$$

where $\beta = \pi/4, \pi/2, 3\pi/4$ for $b_1 = 1, 2, 3$ respectively. In high SNR, using the tight upper bound for the Q-function [33, p.83]

$$Q(x) < \frac{1}{\sqrt{2\pi}x} e^{-x^2/2},\ x > 0, \tag{2.35}$$

we can obtain

$$P_1 = \frac{8}{\gamma}, \quad P_2 = \frac{4}{\gamma}, \quad P_3 = \frac{4}{\gamma} \tag{2.36}$$

where γ is the SNR. Compared with (2.21), $b_1 = 2, 3$ have 3 dB gain over $b_1 = 1$ and $d_{\min}$ is increased by a factor of $\sqrt{2}$ for $b_1 = 2, 3$. Therefore, $d_{\min}$ increases by increasing b_1. For $w > 1$ and other constellations, the SNR gain can be obtained similarly by integrating the pdf of decision variable over the correct decision region. As s_k, the branch factor provides additional SNR gain.

A special case is when $b_1 = |Q|^{w_1}$, $P_b(E_1) = 0$ and $P(\tilde{E}_1^c) = 1$. Therefore, the performance of the GFD is limited by the second window and the diversity order becomes $w_2 + s_1$. Similarly, if $b_i = |Q|^{w_i}$ for $1 \le i < k$, the diversity order of the GFD increases to $w_k + \sum_{i=1}^{k-1} s_i$.

2.2.5 Computational complexity analysis

The computational complexity of an algorithm may be measured in terms of best, worst, or average-case complexity [34]. In practice, the most useful measure proves

to be the worst-case complexity [34]. For example, for the MIMO detectors, the best-case complexity of the SD is simply the complexity of V-BLAST, which is also the best-case complexity of the GFD. The average complexity of the SD derived in [3,6] is exponential in N_T but low in high SNR. The SD has been realized on VLSI circuits in [4,5]. For VLSI circuits implementation, the throughput is limited by the worst-case complexity and the best-case complexity suggests the minimum requirement. Although the likelihood of the worst-case complexity decreases with increasing SNR, the performance depends heavily on such probability. Therefore, we derive the worst-case complexity of the GFD. On the other hand, since we have used both the SD and the computation sharing technique in the GFD, the average complexity of the GFD seems to be difficult to derive analytically. The average complexity determines the energy consumption of a circuit. We determine the average complexity of the GFD in Section 2.2.6 via simulation.

We assume that both R and Q are complex valued. In the first window, the number of flops is

$$C_1(w_1) = \sum_{i=1}^{w_1} |Q|^i[9(w_1+1-i)+4] = 4\frac{|Q|^{w_1+1}-|Q|}{|Q|-1} + 9\frac{|Q|^{w_1+2}-(w_1+1)|Q|^2+w_1|Q|}{(|Q|-1)^2}. \tag{2.37}$$

For the k-th window ($1 < k < K$), the accumulate path metrics for all the combination of the first $t_k = w_{k-1} - s_{k-1}$ variables have been computed in the $(k-1)$-th window. Due to the computation sharing, the number of flops in the k-th window is

$$C_k(w_k) = \sum_{i=t_k+1}^{w_k} |Q|^i[9(w_1+1-i)+4] = C_1(w_k) - C_1(t_k). \tag{2.38}$$

The number of flops in the k-th window is $C_K(w_K) = C_1(w_K)$. Therefore, the total number of flops for the GFD is given by

$$C_{\text{GFD}} = \sum_{k=1}^{K}\prod_{i=1}^{k-1} b_i C_k(w_k). \tag{2.39}$$

For the special case of $w_k = w$, $s_k = s$ and $b_k = 1$, we have

$$C_{\text{GFD}} = C_1(w)+(K-2)(C_1(w)-C_1(w-s))+C_1(w_K) \le KC_1(w)-(K-2)C_1(w-s). \tag{2.40}$$

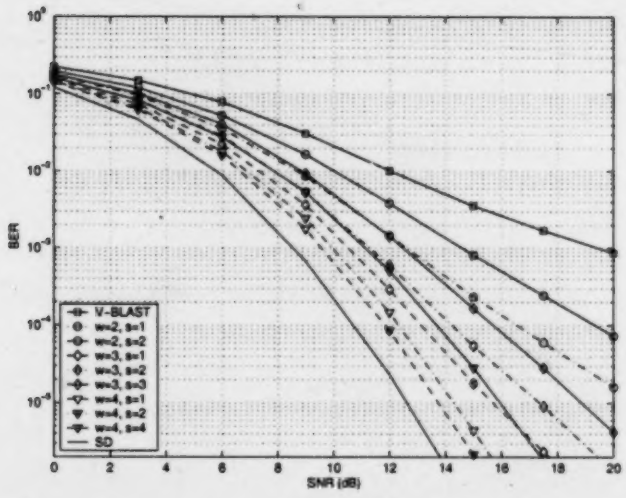

Fig. 2.3. BER comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$.

Interestingly, from (2.37) and (2.40), we find that the worst-case complexity of GFD is exponential in w. Hence, increasing the diversity order of GFD also increases the worst-case complexity exponent. Similarly, as changing step size only changes SNR gain, (2.40) suggests that decreasing step size increases the complexity but does not change the exponent. Therefore, the GFD offers a flexible complexity-performance tradeoff.

2.2.6 Simulation results

We now simulate our GFD for an uncoded MIMO system with 8 transmit and 8 receive antennas over a flat Rayleigh fading channel. The MATLAB V5.3 command "flops" is used to count the number of flops. Only the flops of the search algorithm are counted by ignoring the preprocessing stage. The GFD is compared with V-BLAST and the SD in terms of both performance and complexity. We shall use GFD$[\alpha, \beta, \gamma]$ to denote a GFD with a window size α, step size β, and branch factor γ.

Figs. 2.3 and 2.4 show the BER and average number of flops for different detectors in a BPSK modulated system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$. At

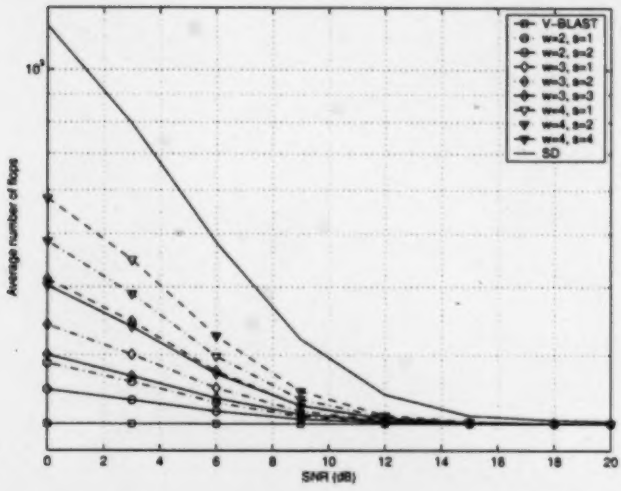

Fig. 2.4. Average complexity comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$.

each stage, we use the SESD [14] and the initial radius is chosen to be infinite. We investigate the effect of window size and step size. Clearly, with different window size, different diversity order is achieved (Fig. 2.3). GFD[2,1,1] has a 3-dB gain over GFD[2,2,1] at BER=10^{-4}. GFD[4,1,1] performs 0.5 dB better than GFD[4,2,1] and 1.5 dB better than GFD[4,4,1] at BER=10^{-5}. Therefore, the SNR gain diminishes with the increasing step size. With different parameter settings, the GFD also has different complexity levels (Fig. 2.4). The complexity of the GFD varies between those of V-BLAST and the SD. In high SNR, all the detectors achieve almost the same average complexity. In low SNR, fixing window size, the average complexity decreases for smaller step size. As well, the average complexity decreases by increasing window size with the same step size. The complexity variation agrees with that of performance, which indicates that better performance results in higher average complexity.

Figs. 2.5 and 2.6 compare the BER and average number of flops for different detectors in a BPSK modulated system. In the GFD, we also set $w_k = w$, $s_k = s$ but $b_k \geq 1$. Instead of using the SESD [14], the initial radius is chosen to be proportional to the noise variance as in [3]. If no solution is found with the initial radius, we

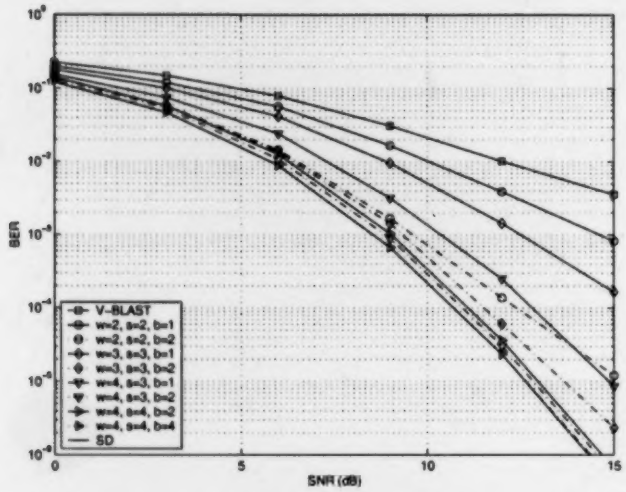

Fig. 2.5. BER comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$.

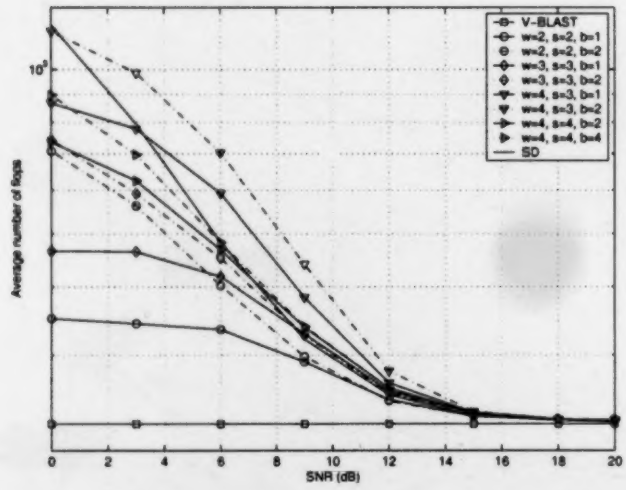

Fig. 2.6. Average complexity comparison of different detectors in an 8×8 BPSK MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$.

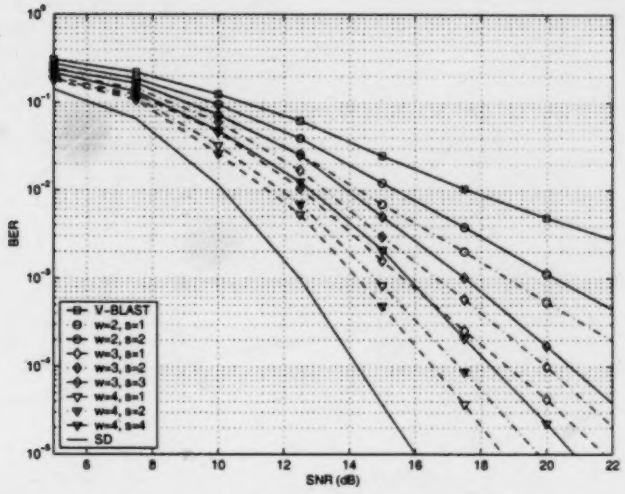

Fig. 2.7. BER comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$.

double the radius until a point is found. Choosing the initial radius in this way is better because the SESD may expand redundant branches in the reduced tree of the GFD and each branch needs high complexity for increasing window size and step size. With the same window size, the diversity order of the GFD is the same but different SNR gains are achieved. GFD[2,2,2] has a 5-dB gain over GFD[2,2,1] at BER= 10^{-3}. The performance gap between GFD[3,3,1] and GFD[2,2,2] reduces to 3 dB at BER= 10^{-3}. GFD[4,4,4] has only a 0.5 dB gain over GFD[4,4,2] at BER= 10^{-4}. GFD[4,4,4] performs close to the SD. Increasing window size, the SNR gain achieved by decreasing the branch factor also diminishes. But with same window size, the SNR gain by increasing branch factor is larger than by increasing step size. With different window size, step size and branch factor, the average complexity is also different. In high SNR, the complexity of all the detectors is the same. However, in low SNR, due to the use of the noise variance initial radius, the average complexity approaches the worst case complexity in low SNR. The complexity of GFD[4,3,2] is higher than that of GFD[4,4,2] because the former forms a 3-level tree while the latter only forms a 2-level tree. In low SNR, GFD[4,4,4] has lower complexity than the SD but with almost the same performance. Therefore, it is an adequate replacement for the SD.

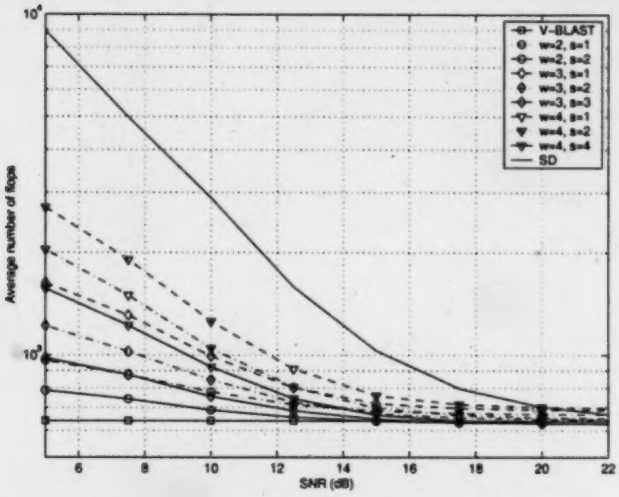

Fig. 2.8. Average complexity comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ and $b_k = 1$.

In Figs. 2.7 and 2.8, we compare different detectors in a 4 QAM system with Gray mapping and $w_k = w$, $s_k = s$ and $b_k = 1$. The SESD [14] is used and the initial radius is chosen to be infinite. Both the performance and complexity results are similar to those of BPSK systems. The SNR gain by increasing step size reduces with the increasing of constellation size. For example, for the BPSK system, GFD[2,1,1] has a 3-dB gain over GFD[2,2,1] at BER=10^{-4}. However, the SNR gain reduces to 2.4 dB for 4QAM at BER=10^{-3}. The complexity difference in high SNR may be due to the number intermediate computations.

Finally, Figs. 2.9 and 2.10 show the BER and average number of flops for different detectors in a 4QAM system with Gray mapping. SESD is not used as in Figs. 2.5 and 2.6. We fix window size and step size and change branch factor to observe its effect on performance. The performance of the B-Chase detector [19] is also evaluated for different list sizes. GFD[2,2,2] has a 4-dB gain over GFD[2,2,1] at a BER of 10^{-3}. GFD[3,3,2] has a 3-dB gain over GFD[3,3,1] at BER=2×10^{-4}. The gain reduces to 2 dB with $w = 3$ and $s = 3$ at BER=10^{-4}. But all the performance gaps become constant in high SNR. GFD[4,4,4] performs close to the SD. The SNR gain achieved by increasing the branch factor also diminishes with the increase of constellation

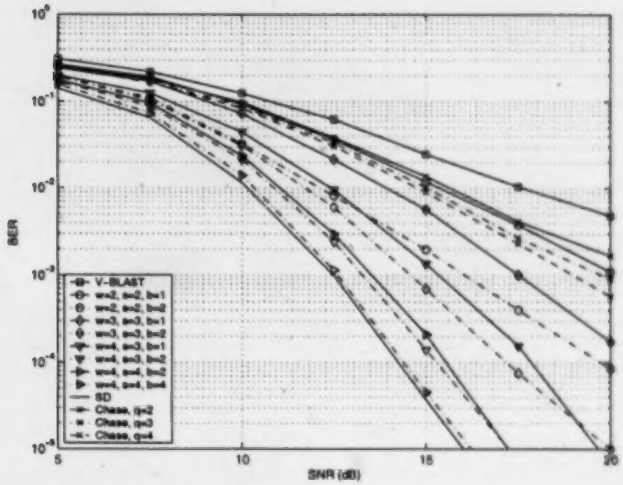

Fig. 2.9. BER comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$.

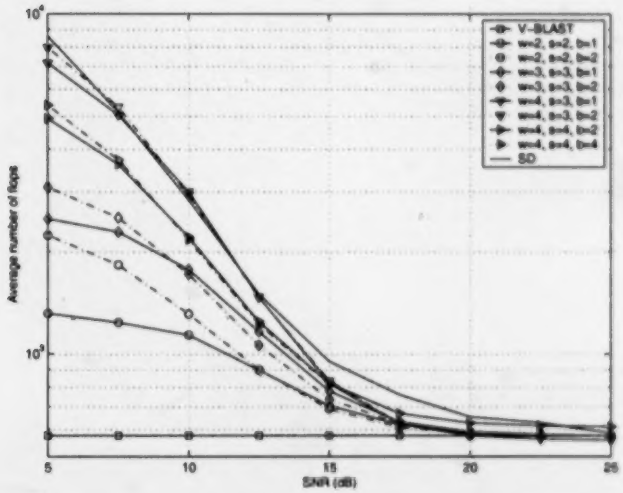

Fig. 2.10. Average complexity comparison of different detectors in an 8×8 4QAM MIMO system. In the GFD, we set $w_k = w$, $s_k = s$ but $b_k \geq 1$.

size. The B-Chase detector performs better than V-BLAST but much worse than the other detectors. The B-Chase detector with $q = 3$ performs better than the B-Chase detector with $q = 2$ as contrary to the prediction by (2.36). Therefore, numerical evaluation of (2.34) is needed to predict the SNR gain. When $q = 4$, the diversity order of B-Chase detector becomes 2, which agrees with our analysis in Section 2.2.4. The complexity results are similar to those in Fig. 2.6. The complexity of all the detectors decreases with the increase of SNR. The complexity difference at high SNR is because of SESD. GFD[4,4,4] has lower complexity than the SD but achieves almost the same performance.

2.3 Constrained Detection for MIMO Systems

2.3.1 Classic linear detectors

The ZF detector relaxes x_i to the whole complex plane and solves

$$\hat{\mathbf{x}}_{\text{ZF}} = \arg\min_{\mathbf{x}\in\mathcal{C}^{N_T}} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2. \tag{2.41}$$

The solution to (2.41) is the well-known least squares (LS) solution and given by

$$\hat{\mathbf{x}}_{\text{ZF}} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{r}. \tag{2.42}$$

The MMSE detector minimizes the mean-square error between the transmitted signals and detected signals $E\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\}$. The MMSE solution is given by

$$\hat{\mathbf{x}}_{\text{MMSE}} = \left(\mathbf{H}^H\mathbf{H} + \sigma_n^2\mathbf{I}_{N_T}\right)^{-1}\mathbf{H}^H\mathbf{r}. \tag{2.43}$$

However, the ZF and MMSE detectors do not guarantee the optimal solution. In general, the solution of an arbitrary linear detector is given by

$$\hat{\mathbf{x}} = \mathbf{G}\mathbf{r} \tag{2.44}$$

where $\mathbf{G}$ is the corresponding filtering matrix for the linear detector.

2.3.2 Real constrained detectors

Even if the constellation $\mathcal{Q}$ is real i.e., BPSK and pulse amplitude modulation (PAM), the ZF and MMSE solutions from (2.42) and (2.43) are usually complex vectors. The imaginary part may cause additional interference since the transmitted vector is real. To impose a real constraint on (2.42) and (2.43), we rewrite (2.1) as

$$\tilde{\mathbf{r}} = \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix} = \begin{bmatrix} \Re\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \Re\{\mathbf{n}\} \\ \Im\{\mathbf{n}\} \end{bmatrix} = \tilde{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{n}}. \tag{2.45}$$

Note that the entries of $\tilde{\mathbf{n}}$ have zero means and variance $\sigma_n^2/2$. The ZF and MMSE detectors for the equivalent system (2.45) can be obtained as

$$\hat{\mathbf{x}}_{\text{R-ZF}} = \left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}\right)^{-1}\tilde{\mathbf{H}}^H\tilde{\mathbf{r}} \tag{2.46}$$

and

$$\hat{\mathbf{x}}_{\text{R-MMSE}} = \left(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}} + \sigma_n^2/2\mathbf{I}_{N_T}\right)^{-1}\tilde{\mathbf{H}}^H\tilde{\mathbf{r}} \tag{2.47}$$

where R-ZF and R-MMSE denote real constrained ZF and MMSE detectors, respectively. The combined detector has tighter relaxation than real constrained detector and modulus constrained detector individually.

2.3.3 Constrained subgroup detectors based on constant-modulus constellations

When $\mathcal{Q}$ is a complex constellation, the constraint on the constellation modulus is exploited. We first consider a constant modulus constellation with unity modulus $|x_i|^2 = 1$. In [24], the CLS relaxes the candidate vectors to be on the hypersphere $\mathbf{x}^H\mathbf{x} = N_T$. To achieve better performance, we must use a tighter relaxation. We partition the vector $\mathbf{x}$ in to g groups and each group forms a subvector $\mathbf{x}_i$ with size s_i, $i = 1, \ldots, g$, where $\sum_{i=1}^{g} s_i = N_T$. We relax each $\mathbf{x}_i$ on a s_i-dimensional hypersphere $\mathbf{x}_i^H\mathbf{x}_i = s_i$. The constrained detector is thus given by

$$\hat{\mathbf{x}}_{\text{CML}} = \underset{\mathbf{x}_1^H\mathbf{x}_1 = s_1, \ldots, \mathbf{x}_g^H\mathbf{x}_g = s_g}{\arg\min} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \tag{2.48}$$

where CML denotes constrained ML detector. The minimization problem (2.48) can be written as

$$\min_{\mathbf{x}} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2$$
$$\text{s.t. } \mathbf{x}_1^H \mathbf{x}_1 = s_1, \ldots, \mathbf{x}_g^H \mathbf{x}_g = s_g. \tag{2.49}$$

The Lagrangian $\mathcal{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_g)$ for this minimization problem is

$$\mathcal{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_g) = \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 + \sum_{i=1}^{g} \lambda_i \left(\mathbf{x}_i^H \mathbf{x}_i - s_i\right). \tag{2.50}$$

Taking partial derivatives with respect to x_i's, the solution for $\mathbf{x}$ can be derived as

$$\hat{\mathbf{x}}(\lambda_1, \ldots, \lambda_g) = \left(\mathbf{H}^H \mathbf{H} + \mathbf{\Lambda}\right)^{-1} \mathbf{H}^H \mathbf{r} \tag{2.51}$$

where $\mathbf{\Lambda}$ is a diagonal matrix and given by

$$\mathbf{\Lambda} = \text{diag}\{\underbrace{\lambda_1, \ldots, \lambda_1}_{s_1}, \ldots, \underbrace{\lambda_g, \ldots, \lambda_g}_{s_g}\}. \tag{2.52}$$

When $g = 1$, there is only one λ_1. Eq. (2.52) reduces to the CLS solution in [24]. When $\lambda_1 = \ldots = \lambda_g = \sigma_n^2$, the CML detector becomes the MMSE detector (2.43).

In order to obtain the CML solution in (2.51), the optimal values for $\lambda_1, \ldots, \lambda_g$ have to be computed so that the constant modulus constraints are fulfilled. Substituting $\hat{\mathbf{x}}(\lambda_1, \ldots, \lambda_g)$ into (2.49), we need to find the zeros of the set of equations

$$F_1(\lambda_1, \ldots, \lambda_g) = \|\hat{\mathbf{x}}_1(\lambda_1, \ldots, \lambda_g)\|^2 - s_1 = 0$$
$$\vdots$$
$$F_g(\lambda_1, \ldots, \lambda_g) = \|\hat{\mathbf{x}}_g(\lambda_1, \ldots, \lambda_g)\|^2 - s_g = 0. \tag{2.53}$$

Note that (2.53) are nonlinear equations but (2.51) has the form of (2.44). Therefore, we also consider the CML detector as a linear detector.

The multidimensional Newton-Raphson root finding method [35] can be used to solve (2.53). It needs to compute the partial derivative of F_i with respect to λ_j, $\partial F_i / \partial \lambda_j$, $1 \leq i, j \leq g$. But $\partial F_i / \partial \lambda_j$ cannot be obtained analytically. Instead, we compute the partial derivatives by finite differences. There are several sets of roots for (2.53) and an initial estimate is needed to guarantee the convergence to desired

root. In CLS where only λ_1 exists, the global minimum is achieved by the maximal real root λ_1^*. But the multidimensional case does not yield the roots minimizing (2.49). Since the MMSE detector (2.43) provides good solution, the initial values for λ_i are chosen as $\lambda_1 = \ldots = \lambda_g = \sigma_n^2$. If the Newton method does not converge after a specified number of iterations, we simply set $\lambda_1 = \ldots = \lambda_g = \sigma_n^2$ or the CML detector outputs the MMSE solution. Simulation results show that the Newton method converges after a few iterations, which thus incurs only a marginal additional complexity to the MMSE detector.

For a non-constant modulus constellation such as QAM, we assume ρ as the largest modulus of the constellation. Similarly, we also partition the vector $\mathbf{x}$ in to g groups. We thus relax each $\mathbf{x}_i$ within a s_i-dimensional hypercube $\mathbf{x}_i^H \mathbf{x}_i \leq \rho^2 s_i$. The CML detector is modified as

$$\hat{\mathbf{x}}_{\text{CML}} = \underset{\mathbf{x}_1^H \mathbf{x}_1 \leq \rho^2 s_1, \ldots, \mathbf{x}_g^H \mathbf{x}_g \leq \rho^2 s_g}{\arg\min} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \tag{2.54}$$

The minimization problem (2.54) can be written as

$$\begin{aligned} &\min_{\mathbf{x}} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \\ &\text{s.t. } \mathbf{x}_1^H \mathbf{x}_1 \leq \rho^2 s_1, \ldots, \mathbf{x}_g^H \mathbf{x}_g \leq \rho^2 s_g. \end{aligned} \tag{2.55}$$

Using the convex duality theorem, the Lagrangian dual function for (2.55) can be expressed as

$$\mathcal{L}(\mathbf{x}, \lambda_1, \ldots, \lambda_g) = \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 + \sum_{i=1}^{g} \lambda_i \left(\mathbf{x}_i^H \mathbf{x}_i - \rho^2 s_i \right). \tag{2.56}$$

Solving (2.56) for $\mathbf{x}$, the solution is the same as (2.51). Substituting it back to (2.56), we obtain

$$\max_{\lambda_1 \geq 0, \ldots, \lambda_g \geq 0} -\mathbf{r}^H \mathbf{H} \left(\mathbf{H}^H \mathbf{H} + \mathbf{\Lambda} \right)^{-1} \mathbf{H}^H \mathbf{r} - \rho^2 \sum_{i=1}^{g} \lambda_i s_i \tag{2.57}$$

where (2.57) is a g-dimensional optimization problem. The simple unconstrained multidimensional gradient descent algorithm can be used to solve (2.57). The partial derivatives are computed by finite differences. If $g = 1$, the CML detector (2.54) reduces to the GMMSE in [23]. Therefore, the CML detector generalizes the GMMSE.

The constrained detectors using constellation modulus can be combined with the real constraint in Section 2.3.2. For real constellations, the CML detectors (2.48) and (2.54) can be directly applied to the equivalent system (2.45). We denote the combined receiver as R-CML.

2.3.4 Chase improvement

An iterative detector can be used to improve the performance of our constrained linear detectors by correcting unreliable decisions of the detector. As before, we partition $\mathbf{x}$ into g groups each with s_i symbols. In each iteration, for $i = 1, \ldots, g$, we fix the other $g-1$ groups and solve

$$\hat{\mathbf{x}}_i = \arg\min_{\mathbf{x}_i \in \mathcal{Q}^{s_i}} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \tag{2.58}$$

where $\mathbf{x}_i$ is the i-th group in $\mathbf{x}$. $\mathbf{x}_i$ is updated to $\hat{\mathbf{x}}_i$. The iteration starts with any solution from a constrained detector and terminates when $\mathbf{x}$ ceases to change during an iteration. Typically, we choose $g = N_T$ and $s_i = 1$. It resembles the chase decoder [36] for soft decoding of linear block codes.

2.3.5 Real decision feedback detectors

For real valued constellations, using the same arguments in Section 2.3.2, V-BLAST detection of (2.45) performs better than that of (2.1) directly and we denote V-BLAST for (2.45) as R-V-BLAST. If $N_T = N_R$ and no permutations are used, the squared-norm of the entries of $\mathbf{R}$ are known to be χ^2 distributed [32], specifically, $|R_{i,i}|^2 \sim \chi^2(2i)$, for $i = 1, \ldots, N_T$ and $|R_{i,j}|^2 \sim \chi^2(2)$, for $j > i$, where $\chi^2(k)$ denotes the chi-squared distribution with k degrees of freedom. Since the performance of V-BLAST is limited by the first detected symbol [18], the diversity order of V-BLAST detection is only one. However, if QR decomposition is performed on the real matrix $\tilde{\mathbf{H}}$ in (2.45), the squared-norm of the entries of $\tilde{\mathbf{R}}$ are also χ^2 distributed but $|R_{i,i}|^2 \sim \chi^2(i + N_T)$, for $i = 1, \ldots, N_T$ and $|R_{i,j}|^2 \sim \chi^2(1)$, for $j > i$. Therefore, it can be readily verified that the diversity order of R-V-BLAST increases to $(N_T + 1)/2$.

For decoupleable complex constellations i.e., QAM, (2.1) can be rewritten as

$$\begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix} = \begin{bmatrix} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{bmatrix} \begin{bmatrix} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{bmatrix} + \begin{bmatrix} \Re\{\mathbf{n}\} \\ \Im\{\mathbf{n}\} \end{bmatrix} \tag{2.59}$$

or

$$\tilde{\mathbf{r}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}. \tag{2.60}$$

In [37], it has been shown that applying V-BLAST to the equivalent real system (2.59) yields an additional performance gain. We can quantify the improvement. In the original system, the diversity order for x_{N_T} is 1. In (2.59), after QR decomposition on $\tilde{\mathbf{H}}$, we find $|R_{i,i}|^2 \sim \chi^2(i)$, for $i = 1, \ldots, 2N_T$. Therefore, the diversity order for $\Re\{x_{N_T}\}$ is $(N_T+1)/2$ and for $\Im\{x_{N_T}\}$ is $1/2$, which may improve the total performance on x_{N_T}.

2.3.6 Constrained ordering decision feedback detectors

The ZF nulling vector $\mathbf{w}_{k_i}$ (1.13) in V-BLAST completely removes the interferences from the other antennas but also amplifies the additive noise. To get a better trade-off between noise enhancement and interference suppression, we use our proposed constrained linear detectors in Sections 2.3.1-2.3.3 instead of the ZF detector in V-BLAST. We replace (1.11) and (1.16) with

$$\mathbf{G}_1 = \left(\mathbf{H}^H\mathbf{H} + \mathbf{\Lambda}\right)^{-1}\mathbf{H}^H \tag{2.61}$$

and

$$\mathbf{G}_{i+1} = \left(\mathbf{H}_{\bar{k}_i}^H\mathbf{H}_{\bar{k}_i} + \mathbf{\Lambda}_i\right)^{-1}\mathbf{H}_{\bar{k}_i}^H \tag{2.62}$$

where $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_i$ can be calculated using (2.53) and (2.57) for constant modulus and non-constant modulus constellations, respectively.

When nulling is performed using CML, interference cannot be removed completely. We thus propose to determine the detection order at each iteration by maximizing the SINR defined as

$$\mathrm{SINR}_j = \frac{|(\mathbf{G}_{i+1}\mathbf{H}_{\bar{k}_i})_{j,j}|^2 E\{|x_j|^2\}}{\sum_{k=1,k\neq j}^{N_T} |(\mathbf{G}_{i+1}\mathbf{H}_{\bar{k}_i})_{j,k}|^2 E\{|x_k|^2\} + \sigma_n^2 \|(\mathbf{G}_{i+1})_j\|^2} \tag{2.63}$$

where $(\mathbf{A})_{i,j}$ is the (i,j)-th entry of matrix $\mathbf{A}$ and $(\mathbf{G}_{i+1})_k$ denotes the k-th row of matrix $\mathbf{G}_{i+1}$. In V-BLAST, (1.17) is replaced by

$$k_{i+1} = \underset{j \notin \{k_1,\ldots,k_i\}}{\arg\min}\ \mathrm{SINR}_j. \tag{2.64}$$

This modified V-BLAST detection is denoted as constrained DFD (CDFD). Note that if $\mathbf{\Lambda} = \sigma_n^2 \mathbf{I}_{N_T}$, CDFD reduces to MMSE-DFD in [22].

2.3.7 Combined constrained linear and decision feedback detectors

We use the ZF-DFD description of V-BLAST detection algorithm. The performance of ZF-DFD is limited by the error propagation of decision feedback. Even though the V-BLAST optimal ordering is employed, the diversity order of V-BLAST detection is only one. This is because the V-BLAST detection is a greedy algorithm. It makes a hard decision only on the "local" metric (1.20) without taking into account its effect on the detection for subsequent symbols. We thus propose to combine the constrained linear detectors in Sections 2.3.1-2.3.3 and ZF-DFD to let the detector make hard decision less greedily. At each iteration, a "global" metric is used to make decision on each symbol, which is obtained by the constrained linear detectors.

In the i-th iteration, we define $\mathbf{R}_i = \mathbf{R}(1:i-1, 1:i-1)$, $\mathbf{r}_i = \mathbf{R}(1:i-1, i)$ and $\mathbf{y}_i = \mathbf{y}(1:i-1)$. For each $x \in \mathcal{Q}$, after cancelling x from $\mathbf{y}$, the soft decisions for the remaining $N_T - i$ symbols can be obtained using the constrained linear detectors as

$$\hat{\mathbf{x}}_i = \left(\mathbf{R}_i^H \mathbf{R}_i + \mathbf{\Lambda}_i\right)^{-1} \mathbf{R}_i^H (\mathbf{y}_i - \mathbf{r}_i x) \tag{2.65}$$

where $\mathbf{x}_i = [x_1, \ldots, x_{i-1}]^T$ and $\mathbf{\Lambda}_i$ can be obtained similarly to (2.53) and (2.57). Since the solution to (2.49) or (2.55) gives a low bound on $\|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2$, the effect of x on the decision metric for the remaining $N_T - i$ symbols can be measured using $\|\mathbf{y}_i - \mathbf{r}_i x - \mathbf{R}_i \hat{\mathbf{x}}_i\|^2$. The global metric for x is defined as

$$\begin{aligned} M_i(x) =& \|\mathbf{y}_i - \mathbf{r}_i x - \mathbf{R}_i \hat{\mathbf{x}}_i\|^2 + |y_i - R_{i,i} x|^2 \\ =& \left\|(\mathbf{I}_{N_T - i} - \mathbf{R}_i (\mathbf{R}_i^H \mathbf{R}_i + \mathbf{\Lambda}_i)^{-1} \mathbf{R}_i^H)(\mathbf{y}_i - \mathbf{r}_i x)\right\|^2 + |y_i - R_{i,i} x|^2. \end{aligned} \tag{2.66}$$

In ZF-DFD, (1.21) is simply replaced by

$$\hat{x}_i = \arg\min_{x \in \mathcal{Q}} M_i(x). \tag{2.67}$$

The resulting detector is denoted CL-DFD. If $\mathbf{\Lambda}_i = \sigma_n^2 \mathbf{I}_{N_T - i}$, (2.65) reduces to MMSE. Though it does not give a low bound on $\|\mathbf{y}_i - \mathbf{r}_i x - \mathbf{R}_i \hat{\mathbf{x}}_i\|^2$, the metric (2.66) also measures the effect of x on the overall metric. Combined MMSE and DFD (CMMSE-DFD) also improves the performance.

2.3.8 Polynomial constrained detector

Due to the finite alphabet nature of $\mathcal{Q}$, each $x_i \in \mathcal{Q}$ satisfies the polynomial equation

$$f(x_i) = \prod_{k=1}^{s} (x_i - q_k) = 0. \tag{2.68}$$

The ML detection problem (2.2) can thus be relaxed as

$$\begin{aligned} &\min_{\mathbf{x} \in \mathcal{C}^{N_T}} \|\mathbf{r} - \mathbf{H}\mathbf{x}\|^2 \\ &\text{s.t. } f(x_i) = \prod_{k=1}^{s} (x_i - q_k) = 0,\ i = 1, \ldots, N_T. \end{aligned} \tag{2.69}$$

Both the objective function and constraints are polynomial in $\mathbf{x}$. For example, for BPSK, $f(x_i) = (x_i - 1)(x_i + 1)$. To avoid the complex operation, Eq. (2.69) can be transformed into a real problem as

$$\begin{aligned} &\min_{\tilde{\mathbf{x}} \in \mathcal{R}^{2N_T}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 \\ &\quad \text{s.t. } g^r(\Re\{x_i\}, \Im\{x_i\}) = 0 \text{ and } g^i(\Re\{x_i\}, \Im\{x_i\}) = 0 \\ &\quad\quad i = 1, \ldots, N_T \end{aligned} \tag{2.70}$$

where $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{H}}$ are defined in (2.59), $g^r(\Re\{x_i\}, \Im\{x_i\}) = \Re\{f(x_i)\}$ and $g^i(\Re\{x_i\}, \Im\{x_i\}) = \Im\{f(x_i)\}$, which are also polynomial in $\Re\{x_i\}$ and $\Im\{x_i\}$. Specifically for decoupleable constellations, i.e., QAM, (2.70) can be simplified as

$$\begin{aligned} &\min_{\tilde{\mathbf{x}} \in \mathcal{R}^{2N_T}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 \\ &\quad \text{s.t. } g(\tilde{x}_i) = 0,\ i = 1, \ldots, 2N_T \end{aligned} \tag{2.71}$$

where $\tilde{x}_i$ is the i-th element of $\tilde{\mathbf{x}}$. For example, for 16-QAM, $g(x) = (x^2-9)(x^2-1) = x^4 - 10x^2 + 9$.

We next show how to apply penalty function method to (2.71). Eq. (2.70) can be solved similarly. The most common penalty function is the one which associates a penalty proportional to the square of the constraints. We thus replace (2.71) by

$$\min_{\tilde{\mathbf{x}} \in \mathcal{R}^{2N_T}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 + \frac{1}{2}c\sum_{i=1}^{2N_T} |g(\tilde{x}_i)|^d \tag{2.72}$$

where the positive scalar c controls the magnitude of the penalty and d controls the acceleration of penalty. If $d = 2$, it reduces to the usual penalty function in [38]. In the following, we choose $d = 2$. Since (2.72) is a polynomial in $\tilde{\mathbf{x}}$, the Hessian matrix of (2.72) can be computed in a closed-form. The well-known Newton or Quasi-Newton method can be used to solve (2.72). It can be initialized with the LS or MMSE solution.

It seems logical to choose a large c to ensure that no constraint is violated. However, large c may lead to numerical difficulties or ill-conditioning, and the search will be trapped by the local minima corresponding to the LS or MMSE initial solution. Consequently, the minimization is started with a relatively small c, and c is increased gradually. A typical value for $c^{(k+1)}/c^{(k)}$ is 5 [38]. If $\tilde{\mathbf{x}}$ is the true solution, $\|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2$ is a chi-square random variable. The initial c can be set to $2\sigma_n^2$.

To overcome the ill-conditioning, the penalty function method can be combined with the Lagrange multipliers. The so-called augmented Lagrangian function [38] is defined as

$$L(\tilde{\mathbf{x}}, \lambda_1, \ldots, \lambda_{2N_T}) = \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\tilde{\mathbf{x}}\|^2 + \sum_{i=1}^{2N_T} \lambda_i g(\tilde{x}_i) + \frac{1}{2}c\sum_{i=1}^{2N_T} |g(\tilde{x}_i)|^2. \tag{2.73}$$

The initial λ_i can be set to $\lambda_i^{(0)} = 0$. After minimizing (2.73), [38] suggests updating λ_i as

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} - c^{(k)} g(\tilde{x}_i^{(k)}) \tag{2.74}$$

where $\tilde{x}_i^{(k)}$ is the estimate of $\tilde{x}_i$ in the k-th iteration.

To avoid the trap of local minima when using the Newton method, we propose a differential-equation algorithm inspired by the classical mechanics to improve the detection performance. Let the function to be minimized in (2.72) be denoted as $F(\mathbf{x})$, where $\tilde{}$ is omitted for brevity. We associate the following second order differential equation with (2.72) [39]:

$$\mu\frac{d^2\mathbf{x}}{dt^2} + \beta(t)\frac{d\mathbf{x}}{dt} + \nabla F(\mathbf{x}) = \mathbf{0}, \tag{2.75}$$

where μ is a positive constant, $\beta(t) > 0$ is a function, and $\nabla F(\mathbf{x})$ is the gradient of $F(\mathbf{x})$.

Eq. (2.75) represents Newton's second law for a particle of mass μ moving in $\mathbb{R}^{N_T}$, subject to the force $-\nabla F(\mathbf{x})$ given by the potential $F(\mathbf{x})$ and the friction $-\beta(t)d\mathbf{x}/dt$ where $\beta(t)$ is the time varying friction coefficient.

Let the initial values for (2.75) be

$$\mathbf{x}(0) = \mathbf{x}_0, \quad \frac{d\mathbf{x}}{dt} = \mathbf{w}_0. \tag{2.76}$$

Typically, $\mathbf{x}_0$ is chosen to be the LS or V-BLAST solution and $\mathbf{w}_0 = \mathbf{0}$. We numerically integrate the differential equation (2.75) with the initial conditions (2.76). Eq. (2.75) can be rewritten as first order equations as

$$\frac{d\mathbf{x}}{dt} = \mathbf{w}, \quad \mu\frac{d\mathbf{w}}{dt} = -\beta(t)\mathbf{w} - \nabla F(\mathbf{x}). \tag{2.77}$$

Applying numerical integration to (2.77), we have

$$\begin{aligned} h_n\mathbf{w}_{n+1} =& \mathbf{x}_{n+1} - \mathbf{x}_n \\ \mu(\mathbf{w}_{n+1} - \mathbf{w}_n) =& -h_n\beta_n\mathbf{w}_n - h_n\nabla F(\mathbf{x}_n) \end{aligned} \tag{2.78}$$

where h_n is the time integration step at time instant t_n. Note that the A-stable linearly implicit method [39] can be used. However, it needs to compute the Jacobian of $F(\mathbf{x})$, which has high complexity.

At each time t_n, we save the current potential $F(\tilde{\mathbf{x}_n})$, kinetic energy $\mu\mathbf{w}_n^2/2$ and the corresponding location $\mathbf{x}_n$. The integration is stopped when the particle stops moving or the maximum kinetic of K time steps is less than a threshold. The point

with the minimum potential on the trajectory is output as the solution for (2.72). If the maximum of $|g(\tilde{x}_i)|^2$ is larger than a threshold, c in (2.72) is increased gradually and search again until the condition is satisfied. Due to the existence of the inertial term or the second order term in (2.75), local minima of $F(\mathbf{x})$ may be overpassed. However, this algorithm does not guarantee the global minimum.

Given the initial value β_0 and $\beta_m > \beta_0$, the friction coefficient β_n is kept constant for the first 10 steps and then is doubled at each step until $2\beta_n > \beta_m$. If $2\beta_n > \beta_m$, β_n is set to β_m and it remains constant during the rest of the integration.

Given the initial value h_0, the value of h_n is updated by a factor of γ. If the total mechanical energy E_n is increased, we choose $\gamma < 1$ or $\gamma > 1$. In the simulation, we choose γ from 1.6 or 0.6.

To keep the total complexity of the differential equations algorithm constant, we can set the maximum number of time integration steps $N_{\max}$ in the algorithm. The integration stops after reaching $N_{\max}$.

2.3.9 Simulation results

The error rates of our proposed constrained detectors are simulated for a MIMO system with 8 transmit and 8 receive antennas over a flat Rayleigh fading channel. We assume the receiver has perfect channel state information (CSI) and noise variance. We use notation Chase-X to denote the combination of the detector X and iterative correction in Section 2.3.4.

Fig. 2.11 show the BER performance of different constrained linear detectors in a BPSK modulated system. We compare our detectors with SD [2] and the CLS detector [24]. When all the linear detectors are applied to the complex system (2.1), the CLS and CML perform close to MMSE. In high SNR, CML with $g = 8$ performs better than MMSE. But the performance of all of them is inferior to the ML performance achieved by SD. When the detectors are applied to the real system (2.45), the performance of all the detectors improve. At BER=10^{-3}, R-MMSE has a 0.5-dB gain over R-CLS. Both R-CML with $g = 4$ and $g = 8$ perform better than R-MMSE.

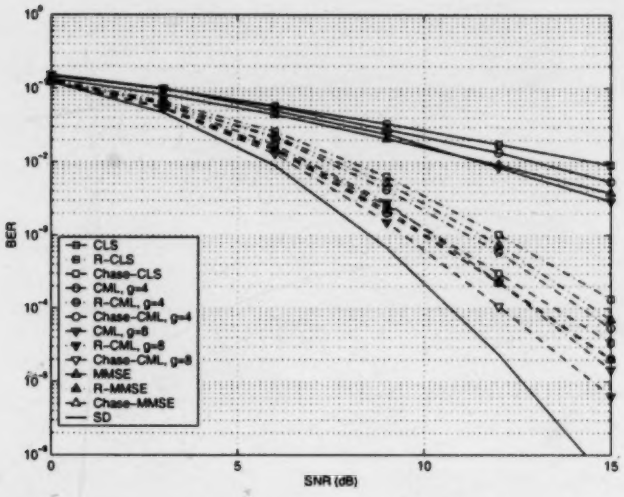

Fig. 2.11. Performance comparison of constrained linear detectors in an 8×8 MIMO system with BPSK.

They have 0.3 dB and 2 dB gain over R-MMSE, respectively. After employing the iterative improvement to all the detectors, R-MMSE, R-CLS and R-CML with $g = 4$ have 2 dB, 1.8 dB and 1.5 dB gains at BER=10^{-3}. The detector R-CML with $g = 8$ improves by 1 dB at BER=10^{-4}.

The BER of GMMSE [23] and different constrained linear detectors for 16QAM is shown in Fig. 2.12. GMMSE performs worst among all the detectors. CML with $g = 4$ has a 0.8 dB loss over MMSE at BER=10^{-3}. In low SNR, CML with $g = 8$ performs better than MMSE, but they perform identically in high SNR. With the Chase iterative improvement, R-MMSE, R-CLS, R-CML with $g = 4$ and R-CML with $g = 8$ have 2 dB, 1.8 dB 1 dB and 1.2 dB gains at BER=10^{-2}, respectively. Since the group-wise hypercube constraint (2.49) is loose, the resulting performance improvement is marginal. Tighter constraints are needed for high order QAM constellations.

Fig. 2.13 compares the BER of DFD and real DFD with different constrained ordering schemes. BPSK modulation is used. The performance of V-BLAST and SD is also shown in Fig. 2.13. We observe a dramatic performance improvement even for the constrained DFDs on the complex model. At BER=10^{-2}, the CDFDs and

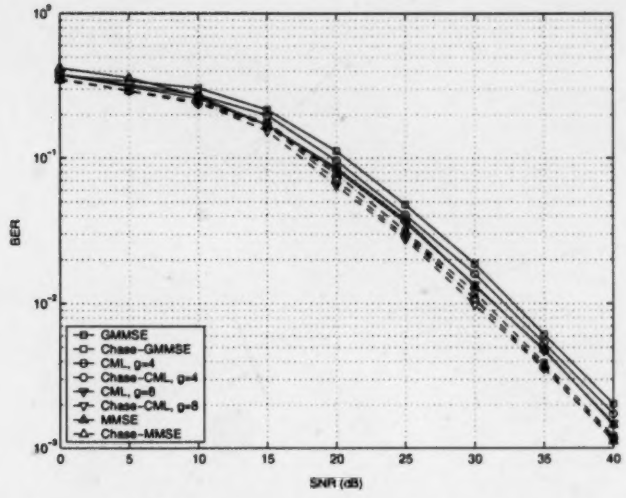

Fig. 2.12. Complexity of constrained linear detectors in an 8×8 MIMO system with 16QAM.

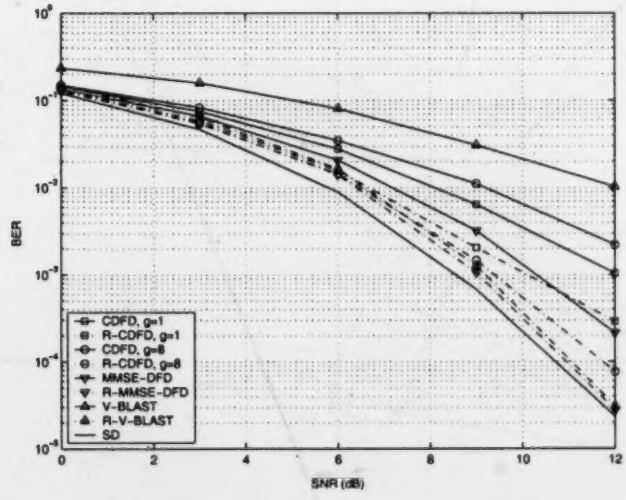

Fig. 2.13. Complexity of constrained ordering decision feedback detectors in an 8×8 MIMO system with BPSK.

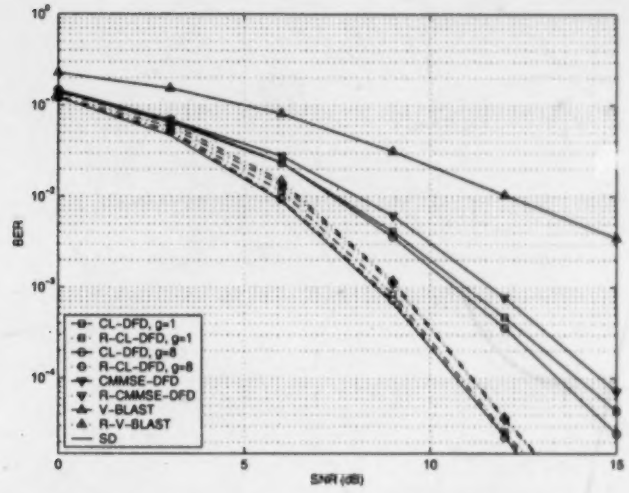

Fig. 2.14. Complexity of combined linear and decision feedback detectors in an 8×8 BPSK MIMO system.

MMSE-DFD have more 3 dB gain over V-BLAST. Therefore, the CDFD and the MMSE-DFD have a smaller noise enhancement compared to the ZF-DFD. When the real constraint is imposed, R-V-BLAST and R-MMSE-BLAST perform close to SD at high SNR. They both perform only about 0.2 dB worse than SD at BER=10^{-4}. The gap between R-CDFD with $g = 8$ and R-V-BLAST is 0.7 dB at BER=10^{-4}. Since the diversity order of R-V-BLAST is $(N_T + 1)/2$, it performs well and the performance improvement by using R-MMSE-BLAST is small.

We then present the results for combined linear and decision feedback detectors in a BPSK system (Fig. 2.14). Clearly, our proposed CL-DFDs significantly improve the performance, indicating their ability to mitigate error propagation. At BER=10^{-2}, the CL-DFD with $g = 1$ has a 4-dB gain over V-BLAST. The CL-DFD with $g = 1$ performs worse than the CMMSE-DFD. But the CL-DFD with $g = 8$ has better performance than the CMMSE-DFD. At BER=10^{-4}, the CL-DFD with $g = 8$ performs 0.5 dB better than the CMMSE-DFD. When the real constraint is applied, all the detectors perform close to SD at high SNR. R-CL-DFD with $g = 8$ can almost achieve the ML performance. However, the performance gain using our R-CL-DFDs decreases compared to the complex case. For non-constant modulus constellations,

the performance gain by using CL-DFD and MMSE-DFD is also significant. More results are omitted for brevity.

Finally, we show the performance of our polynomial constrained detector, which is denoted as "PCD". The detector without maximum number of time steps constraint is denoted as "PCD-Op" or it is denoted as "PCD-X", where X is $N_{\max}$. The LS detector and SD are used as benchmark detectors. The initial value for $\mathbf{x}_0$ is chosen as the LS solution.

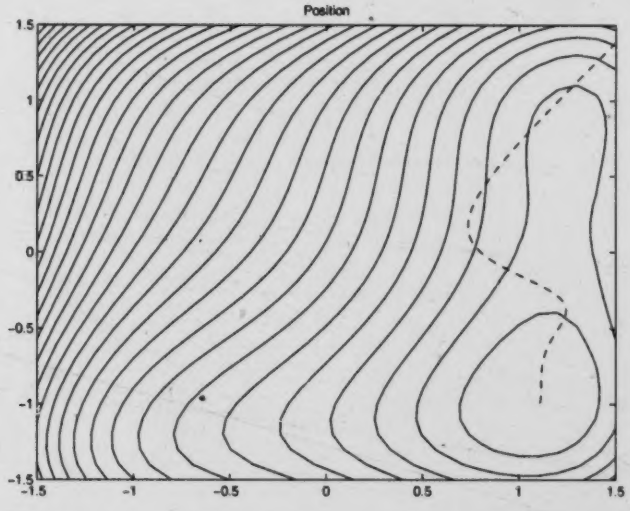

Fig. 2.15. The trajectory of a particle for a 2×2 MIMO system with BPSK and 5 dB.

We first show a simple example of 2×2 BPSK system at SNR=5 dB. Fig. 2.15 shows the trajectory of the particle (dash line) on a contour graph. The initial point is set to $\mathbf{x}_0 = [1.5, 1.5]^T$. Clearly, we can see the particle is not trapped by the local minima around $[1, 1]$ and it stops at the global minima around $[1, -1]$. Fig. 2.16 shows the kinetic, potential and the total mechanical energy as a function of the time steps. From the potential curve, we find the particle overpasses a local minima at time step 20. In the end, the kinetic energy becomes zero as the particle reaches the global minima.

Fig. 2.17 shows the BER performance of different detectors in a BPSK modulated

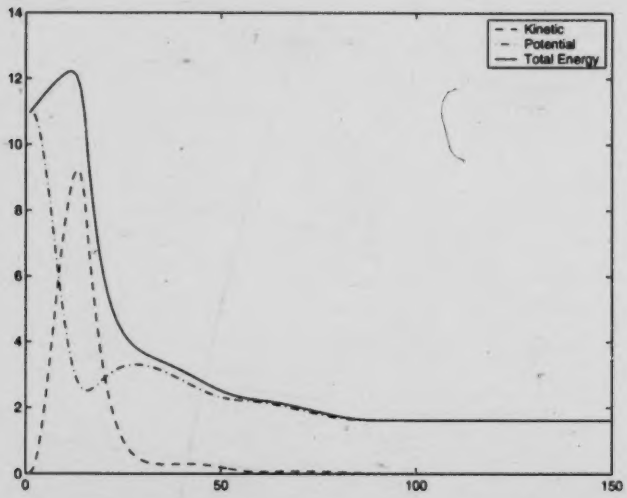

Fig. 2.16. The energy of a particle as a function of time steps for a 2×2 MIMO system with BPSK and 5 dB.

system with 8 transmit and 8 receive antennas. Our PCD has significant performance gain over V-BLAST and LS. At BER = 10^{-2}, PCD-Op has 4 dB gain over V-BLAST and the performance loss over ML is only 2.2 dB. When PCD-40 is used to achieve constant complexity, the performance loss over PCD-Op is less than 0.1 dB at BER = 10^{-2}. However, the complexity of PCD-40 is roughly 15% of that of ML search. Given a vector multiplication unit such as Matlab, the computation time of PCD is less than that of SD, even though the total number of flops of PCD is more than that of SD.

The symbol error rate (SER) of different detectors for a 4×4 system with 16QAM is shown in Fig. 2.18. Our PCD-Op still has a 5.8 dB gain over LS at SER = $\times 10^{-1}$. However, the performance improvement is reduced compared to that of BPSK. In high SNR, the performance gap between PCD-Op and SD is large. However, the complexity of our PCD is only 1% of that of ML search. The diversity order of PCD-Op appears to be one. The performance gain by using our PCD decreases since the differential equations algorithm may also be trapped by local minima especially when the constellation size is large.

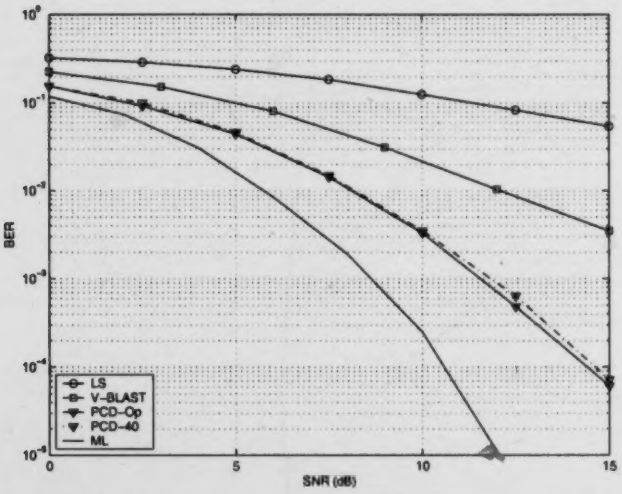

Fig. 2.17. Performance comparison of different detectors in an 8×8 MIMO system with BPSK.

2.4 Multistage Sphere Decoder for High Order Constellations

2.4.1 Multistage sphere decoding algorithm

Fig. 2.19 shows the block disgram of the multistage sphere decoder for 16QAM systems. We denote the sphere decoder in Section 1.2.2 as single stage sphere decoder (SSD), which computes all x's that lie within a sphere of given radius. We use the formulation (2.8).

For brevity, we only show how to apply the multistage sphere decoder to 16QAM and a more general algorithm is given later. An arbitrary 16QAM vector $\mathbf{x}$ can be uniquely expressed as $\mathbf{x} = \sqrt{2}\mathbf{x}_1 + \sqrt{2}/2\mathbf{x}_2$, where $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q}_4^{N_T}$. Similarly, let the true transmit vector be $\mathbf{x}^* = \sqrt{2}\mathbf{x}_1^* + \sqrt{2}/2\mathbf{x}_2^*$. The problem of detecting $\hat{\mathbf{x}}$ (2.2) is equivalent to detecting two 4QAM component vectors as follows:

$$[\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2] = \arg\min_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{Q}_4^{N_T}} \left\| \mathbf{y} - \mathbf{R}(\sqrt{2}\mathbf{x}_1 + \frac{\sqrt{2}}{2}\mathbf{x}_2) \right\|^2. \tag{2.79}$$

To begin, we need an initial approximation to the true signal $\mathbf{x}^*$. Let this be $\tilde{\mathbf{x}} =$

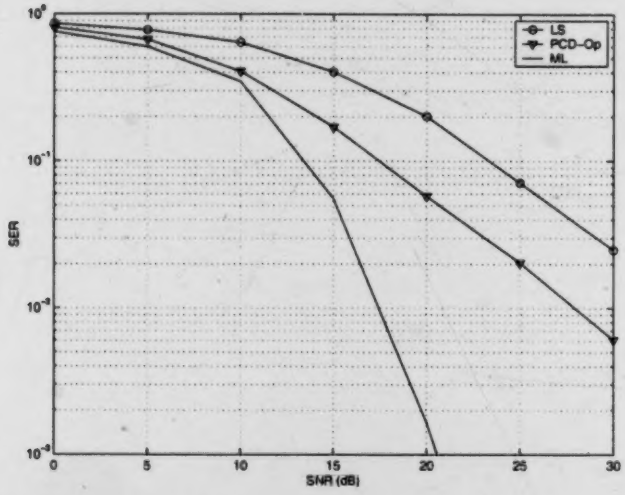

Fig. 2.18. Performance comparison of different detectors in an 4×4 MIMO system with 16QAM.

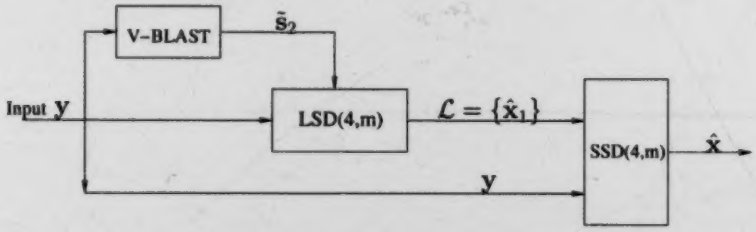

Fig. 2.19. The block diagram of the multistage sphere decoder for 16QAM systems.

$\sqrt{2}\tilde{\mathbf{x}}_1 + \sqrt{2}/2\tilde{\mathbf{x}}_2$. Using this, we do a partial interference cancelation as $\mathbf{y}_2 = \mathbf{y} - \sqrt{2}/2\mathbf{R}\tilde{\mathbf{x}}_2$ in the first stage. Note that we cancel $\tilde{\mathbf{x}}_2$ first since any errors in $\tilde{\mathbf{x}}_2$ will be attenuated by $\sqrt{2}$ (see Eq. (2.79)). Whereas any errors in $\tilde{\mathbf{x}}_1$ will be magnified by $\sqrt{2}$. If $\mathbf{x}_2 = \mathbf{x}_2^*$, $\mathbf{y}_2$ is clearly sufficient to detect $\mathbf{x}_1^*$. We search $\mathbf{x}_1$ that minimizes

$$\|\mathbf{y}_2 - \sqrt{2}\mathbf{R}\mathbf{x}_1\|^2. \tag{2.80}$$

However, $\tilde{\mathbf{x}}_2 \neq \mathbf{x}_2^*$ in general, and minimizing (2.80) will likely give a wrong estimate. Therefore, we use a LSD [30] to generate a list $\mathcal{L}$ of the N_{cand} candidates $\hat{\mathbf{x}}_1$ that make (2.80) smallest. The list size is between 4^{N_T} and 1, and is proportional to the probability that the true solution $\mathbf{x}_1^*$ falls in the list. With a properly chosen radius d, we can obtain $\mathcal{L}$ with N_{cand} candidates on average. To obtain a typical value of d,

we note that for true $\mathbf{x}_1^*$

$$\|\mathbf{y}_2 - \sqrt{2}\mathbf{R}\mathbf{x}_1^*\|^2 = \left\| \frac{\sqrt{2}}{2}\mathbf{R}(\mathbf{x}_2^* - \tilde{\mathbf{x}}_2) + \mathbf{n} \right\|^2 \tag{2.81}$$

where $\mathbf{n}$ is the additive Gaussian noise vector with variance σ_n^2. Since $\tilde{\mathbf{x}}_2$ is correlated with $\mathbf{R}$ and $\mathbf{n}$, (2.81) cannot be treated as a chi-square random variable with $2N_R$ degrees of freedom. The expected value of this random variable is denoted by E, which can be obtained via simulation. As in [30], one possible choice of radius is $d^2 = kE$, where k is chosen so that the average length of the list is N_{cand}. For typical values of σ_n^2 and σ_h^2, d^2 corresponding to N_{cand} can be obtained from simulation and can be stored in memory for practical use.

In the second stage, for each candidate $\hat{\mathbf{x}}_1 \in \mathcal{L}$, the SSD$(4, N_T)$ solves

$$\hat{\mathbf{x}}_2 = \arg \min_{\mathbf{x}_2 \in \mathcal{Q}_4^{N_T}} \left\| \mathbf{y}_1 - \frac{\sqrt{2}}{2}\mathbf{R}\mathbf{x}_2 \right\|^2 \tag{2.82}$$

where $\mathbf{y}_1 = \mathbf{y} - \sqrt{2}\mathbf{R}\hat{\mathbf{x}}_1$. This process provides N_{cand} pairs of $[\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2]$ and the best among them is selected as the output. Each time a $\hat{\mathbf{x}}_2$ for (2.82) is found with a $\hat{\mathbf{x}}_1$, the search radius of the second stage is updated if $\|\mathbf{y}_1 - \sqrt{2}/2\mathbf{R}\mathbf{x}_2\|$ is less than the current radius.

Remarks:

- Our proposed multistage sphere decoder consists of an LSD(4, N_T) and an SSD(4, N_T), which computes N_{cand} points of 4QAM vector pairs. Is our approach less complex than an SSD(16, N_T)? That depends on the difference of the complexities of SSD(4, N_T) and SSD(16, N_T). For example, in the high SNR region the difference of complexity is small, and our proposed multistage sphere decoder is not more efficient than the SSD$(16, N_T)$, while in the low SNR region the difference of complexity can be large, our proposed algorithm is much more efficient.

- The parameter N_{cand} gives a tradeoff between complexity and performance. When N_{cand} is small, the probability that $\mathbf{x}_1^* \in \mathcal{L}$ is low and the BER will increase, while the complexity will decrease.

- The initial estimate $\tilde{\mathbf{x}}_2$ can be obtained via ZF, MMSE or V-BLAST.
- The performance of our multistage sphere decoder may be further improved by using $\hat{\mathbf{x}}_2$ from the second stage SD to recompute the first stage output $\hat{\mathbf{x}}_1$. This will result in an iterative multistage sphere decoder and will increase computational complexity.

Our proposed multistage sphere decoder can be generalized to other constellations. For example. a 64QAM vector $\mathbf{x}$ can be uniquely expressed as $\mathbf{x} = 2\sqrt{2}\mathbf{x}_1 + \sqrt{2}\mathbf{x}_2 + \sqrt{2}/2\mathbf{x}_3$, where $\mathbf{x}_i \in \mathcal{Q}_4^{N_T}$, $i = 1, 2, 3$. Therefore. the multistage sphere decoder will have three stages. In the first stage, an LSD is used to generate a list of $\mathbf{x}_1$. Another LSD is used to generate a list of $\mathbf{x}_2$ in the second stage. In the last stage, an SSD is used to obtain the solution. Similar to (2.5), the 4^q-QAM constellation can be represented as a weighted sum of q QPSK constellations [27]. That is, for $s \in M$-QAM and $s_i \in$ QPSK, $0 \leq i < q$, we have

$$s = \sum_{i=0}^{q-1} 2^i \left(\frac{\sqrt{2}}{2}\right) s_i. \tag{2.83}$$

Hence the multistage sphere decoder for 4^q-QAM system has q stages. For 2^q-PSK, the multistage sphere decoder has q stages and consists of $q-1$ LSD's. For brevity, we do not give the algorithm in detail.

The complexity of a q-stage multistage sphere decoder $C_{\text{MSD}} < (\prod_{i=1}^{q-1} N_{\text{cand},i} + q - 1)C_{\text{SD}}$, where $N_{\text{cand},i}$ is the list size of the i-th stage and C_{SD} is the complexity of the SSD with 4QAM. Simulation results show that the multistage sphere decoder achieves increased complexity savings for large constellations.

The multistage sphere decoder algorithm for 4^q-QAM can be summarized in **Algorithm 2**.

2.4.2 Simulation results

We now compare the multistage sphere decoder (MSD) with the SSD for a 16QAM, uncoded MIMO system with 4 transmit and 4 receive antennas over a flat Rayleigh

input : $\mathbf{y}$, $\mathbf{R}$, d.
output: The suboptimal $\mathbf{x}$.

Compute the ZF solution $\tilde{\mathbf{x}}$ and decomposed it to $\tilde{\mathbf{x}} = \sum_{i=1}^{q} \alpha_i \hat{\mathbf{x}}_i$, where $\hat{\mathbf{x}}_i \in \mathcal{Q}_4^{N_T}$;
for $k \leftarrow 1$ *to* $q-1$ **do**
    $\mathbf{y}_k = \mathbf{y} - \sum_{i \neq k} \alpha_i \mathbf{R} \hat{\mathbf{x}}_i$;
    Solve $\|\mathbf{y}_k - \alpha_k \mathbf{R} \hat{\mathbf{x}}_k\|^2 < r^2$ with an LSD and insert each $\hat{\mathbf{x}}_k$ into a list $\mathcal{L}_k$;
end
for *each* $(q-1)$ *candidates* $[\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{q-1}]$ **do**
    Solve

$$\hat{\mathbf{x}}_q = \arg\min_{\mathbf{x}_q \in \mathcal{Q}_4^{N_T}} \|\mathbf{y}_q - \alpha_q \mathbf{R} \mathbf{x}_q\|^2 \tag{2.84}$$

    with an SSD, where $\mathbf{y}_q = \mathbf{y} - \sum_{i=1}^{q-1} \alpha_i \mathbf{R} \hat{\mathbf{x}}_i$;
end
Find the best q-tuple $[\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_q]$ and output $\hat{\mathbf{x}} = \sum_{i=1}^{q} \alpha_i \hat{\mathbf{x}}_i$;

Algorithm 2: Multistage Sphere Decoding Algorithm

fading channel. Only the flops of the search algorithm are counted without accounting for the preprocessing stage. The initial radius d is chosen according to the noise variance. Both the SSD and the multistage sphere decoder use the SESD [14]. The initial detection uses the ZF-VBLAST.

Fig. 2.20 compares the BER of the SSD with that of the multistage sphere decoder as a function of the number of candidates in the first stage N_{cand}. As N_{cand} increases, the multistage sphere decoder performs close to the SSD. As N_{cand} varies, its performance varies between those of V-BLAST and SSD. The complexity of the multistage sphere decoder increases as N_{cand} increases (Fig. 2.21) and it is lower than that of the SSD's when the SNR is below a threshold. For instance, when $N_{\text{cand}} = 10$, this complexity crossover point is 17 dB. Fig. 2.21 also shows that the complexity of the multistage sphere decoder is almost constant with specific N_{cand}, suggesting that its complexity is polynomial for the whole SNR range in contrast to the conventional SD. The major drawback of our multistage sphere decoder is that the complexity needed to achieve near ML performance increases with increasing SNR. Thus our multistage sphere decoder is suitable for the low SNR region, where it can be combined with an outer code to achieve low BER.

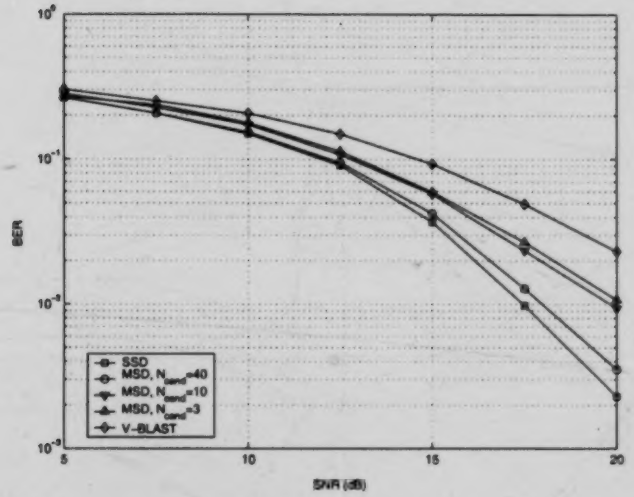

Fig. 2.20. BER comparison with different N_{cand} for a 16QAM MIMO system with 4 transmit and 4 receive antennas.

2.5 Conclusion

In this chapter, we have proposed a unified framework for MIMO detection. Our GFD generalizes the classical feedback decoding for convolutional codes. The GFD varies between SD and V-BLAST in terms of both complexity and performance. By deriving the union bound for the symbol error probability of the GFD, we showed that it achieves an arbitrary diversity order between 1 and N and different SNR gains. We also established the connection between MIMO detectors and tree search algorithms. Moreover, a shared computation technique was proposed to further reduce the complexity. We also considered the relaxation approach to the MIMO detection problem. A class of constrained linear detectors and a class of constrained decision feedback detectors were developed. A polynomial constrained detector was also proposed and solved using penalty function and differential equations. For high order constellation applications, we derived a multistage sphere decoding algorithm, which exploits that many higher-order signal constellations can naturally be decomposed into several lower-order constellations.

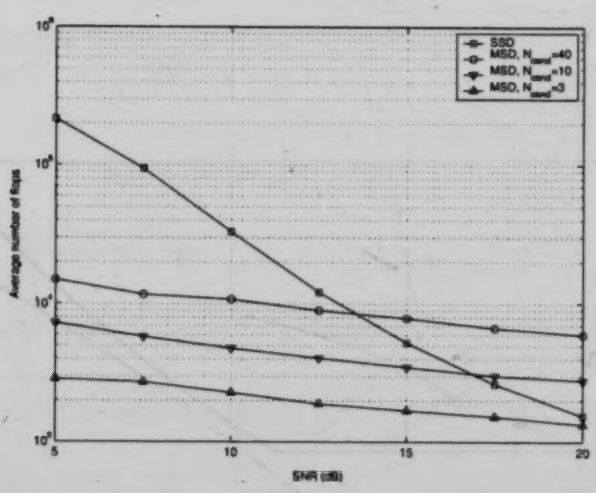

Fig. 2.21. Complexity comparison with different N_{cand} for a 16QAM MIMO system with 4 transmit and 4 receive antennas.

Chapter 3

Blind Decoding for Orthogonal Space-Time Block Codes

This chapter considers the efficient joint channel estimation and decoding of orthogonal space-time block coded (OSTBC) multiple-input multiple-output (MIMO) systems over flat Rayleigh fading channels. This chapter is organized as follows. Section 3.1 introduces the system model and OSTBC. Section 3.2 derives a general maximum likelihood (ML) blind decoder. Efficient solution for the decoder is also discussed. In Section 3.3, the totally blind decoder and a superimposed training scheme are presented. Section 3.4 gives the minimum mean-square-error (MMSE) channel estimator. Section 3.5 gives numerical results, and Section 3.6 concludes this chapter.

3.1 Introduction

3.1.1 Background

Space-time block coding (STBC) with orthogonal designs [7, 40] is one of the major techniques for multi-antenna wireless systems used to effectively utilize diversity gains. OSTC achieves full transmit diversity and is amenable to simple ML decoding if the

channel state information (CSI) is known at the receiver. However, a multiple antenna channel is difficult to estimate and may vary rapidly due to the users' mobility. CSI estimation using pilot symbols will reduce the effective data rate. These factors have motivated the blind detection for OSTBC.

In [41], a suboptimal blind detector (cyclic detector) has been proposed to approximate blind ML decoding of OSTBC, which does not guarantee global optimization. A subspace based decoder was proposed in [42], which does not show ML performance. Recently, an efficient blind ML decoder using semi-definite relaxation (SDR) was given in [43]. This SDR-ML decoder provides a substantially better bit error rate (BER) than the former blind decoders. However, it is applicable only for binary phase shift keying (BPSK), and also needs pilot symbols to solve the phase ambiguity. All of these papers assume that the channel remains constant during several blocks. In practice, however, the mobile channel is time-varying due to the users' mobility.

In this chapter, we derive a general decision rule for the ML blind OSTC decoding in a quasi static (QS) fading channel instead of assuming a constant channel over several blocks. That is, we assume the channel remains constant for one block and allow it to vary from block to block. However, for a more realistic study, we allow the channel to fade continuously in the simulation. Using the linear dispersion property of OSTBC, we show that the decision rule is a quadratic minimization problem. Instead of using exhaustive search, we solve it using sphere decoding (SD) [2] or V-BLAST [1]. To solve the inherent phase ambiguity, pilot symbols may be transmitted. To improve the bandwidth efficiency, a novel approach for totally blind decoding without any pilots is presented using two different PSK constellations. A superimposed training scheme is also presented. Moreover, we give a minimum mean-square-error (MMSE) channel estimator and derive the Cramér-Rao bound (CRB). Power allocation is also discussed in the end.

3.1.2 System model

We consider a MIMO system with N_T transmit and N_R receive antennas. Each block of transmitted symbols has T time slots and time interval T_B. The symbols transmitted during the nth block are denoted by the $T \times N_T$ matrix $\mathbf{S}[n] = [s_{t,i}[n]]$, $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, N_T$, where $s_{t,i}[n]$ is transmitted by the ith antenna in the $t+(n-1)T$-th time slot. For OSTBC, P symbols $\mathbf{x}[n] = [x_1[n], x_2[n], \ldots, x_P[n]]^T$ are transmitted in the nth block with the same average power $E_s = E\{|x_p[n]|^2\}$. $\mathbf{S}[n]$ is formulated using $\mathbf{x}[n]$, and has the property

$$\mathbf{S}^H[n]\mathbf{S}[n] = c\sum_{p=1}^{P}|x_p[n]|^2\mathbf{I}_{N_T}, \tag{3.1}$$

where $c = 1/r$, and $r = P/T$ is the rate of the code. The orthogonal property enables simple symbol by symbol decoding at the receiver. For Alamouti code [40] or the $\mathcal{G}_2$ code of [7], $N_T = 2$, $P = 2$, $T = 2$, $c = 1$ and $\mathbf{S}[n]$ is given by

$$\mathbf{S}[n] = \begin{pmatrix} x_1[n] & x_2[n] \\ -x_2^*[n] & x_1^*[n] \end{pmatrix}. \tag{3.2}$$

For the $\mathcal{G}_3$ and $\mathcal{G}_4$ codes of [7], $c = 2$.

The OSTBC can be alternately represented as [44]

$$\mathbf{S}[n] = \sum_{p=1}^{P}\left(\alpha_p[n]\mathbf{A}_p + j\beta_p[n]\mathbf{B}_p\right), \tag{3.3}$$

where $x_p[n] = \alpha_p[n] + j\beta_p[n]$ and $\mathbf{A}_q$, $\mathbf{B}_q$ are called dispersion matrices [44], which are specified by the OSTBC. Eq. 3.3 is the linear dispersion property of OSTBC. For Alamouti code, we have

$$\begin{aligned} \mathbf{A}_1 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\ \mathbf{B}_1 &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \end{aligned} \tag{3.4}$$

We consider a frequency-flat Rayleigh fading MIMO channel and assume a rich scattering environment. The received signal at the jth receive antenna at time slot t

in the nth block can be written as

$$r_{t,j}[n] = \sum_{i=1}^{N_T} h_{i,j}[n]s_{t,i}[n] + w_{t,j}[n] \tag{3.5}$$

where $h_{i,j}[n]$ denotes the path gain from the ith transmit antenna to the jth receive antenna and $w_{t,j}[n]$ is the complex additive white Gaussian noise at the jth receive antenna with zero mean and variance σ_n^2. The fading channel is assumed to be QS, i.e., channel variations within each block are negligible. All path gains are statistically independent ($E\{h_{i,j}[n]h^*_{i',j'}[n]\} = 0$) and have the same time correlation function $R_h(\tau)$. Therefore, $h_{i,j}[n]$ has correlation $R_h[m] = R_h(mT_B)$. Typically, when Clarke's model [45] is used, $R_h[m]$ is given by

$$R_h[m] = E\{h_{i,j}[n]h^*_{i,j}[n+m]\} = \sigma_h^2 J_0(2\pi m f_d T_B) \tag{3.6}$$

where σ_h^2 denotes the power of the path gain, $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, and f_d is the Doppler frequency due to users' mobility. Note that the QS condition is met when $f_d T_B < 0.03$. Eq. (3.5) can be written in matrix form as

$$\mathbf{R}[n] = \mathbf{S}[n]\mathbf{H}[n] + \mathbf{W}[n] \tag{3.7}$$

where $\mathbf{R}[n] = [r_{t,j}[n]]$ is the $T \times N_R$ receive matrix, $\mathbf{H}[n] = [h_{i,j}[n]]$ is the $N_T \times N_R$ channel matrix, and $\mathbf{W}[n] = [w_{t,j}[n]]$ is the $T \times N_R$ noise matrix. The code transmission format and channel are shown in Fig. 3.1.

3.2 Maximum-Likelihood Blind Decoder

This section derives a new general ML metric for blind decoding. The blind decoder decodes the transmitted symbols in N consecutive blocks. We consider the sequence from $n = k+1$ to $n = k+N$. Let $\bar{\mathbf{R}}[k] = [\mathbf{R}^H[k+1], \mathbf{R}^H[k+2], \ldots, \mathbf{R}^H[k+N]]^H$ and $\bar{\mathbf{S}}[k] = [\mathbf{S}^H[k+1], \mathbf{S}^H[k+2], \ldots, \mathbf{S}^H[k+N]]^H$. The ML decision rule for the sequence $\bar{\mathbf{S}}[k]$ can be expressed as

$$\bar{\mathbf{S}}[k] = \arg\max_{\bar{\mathbf{S}}[k]} f(\bar{\mathbf{R}}[k]|\bar{\mathbf{S}}[k]), \tag{3.8}$$

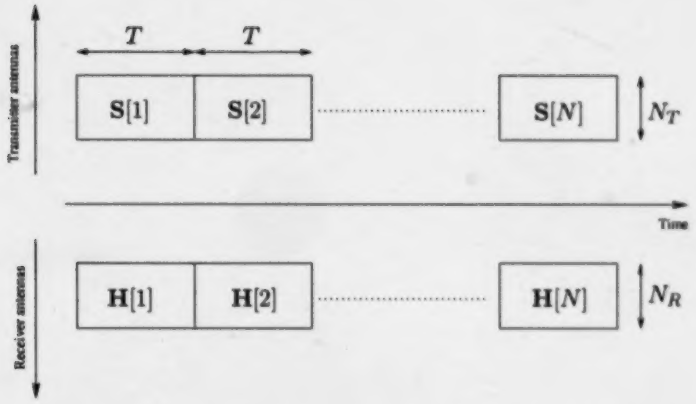

Fig. 3.1. The transmission diagram of a space time block coded system.

where $f(a|b)$ is the probability density function (pdf) of a conditioned on b. The conditional pdf (3.8) can be calculated by averaging the pdf $f(\bar{\mathbf{R}}[k]|\bar{\mathbf{S}}[k], \bar{\mathbf{H}}[k])$ with respect to the channel matrix $\bar{\mathbf{H}}[k]$, which results in

$$f(\bar{\mathbf{R}}[k]|\bar{\mathbf{S}}[k]) = \frac{1}{(\pi^{NN_T}\det(\mathbf{C}_R[k]))^{N_R}} \exp\left(-\mathrm{tr}\left(\bar{\mathbf{R}}^H[k]\mathbf{C}_R^{-1}[k]\bar{\mathbf{R}}[k]\right)\right), \tag{3.9}$$

where $\bar{\mathbf{H}} = [\mathbf{H}^T[k+1], \mathbf{H}^T[k+2], \ldots, \mathbf{H}^T[k+N]]^T$ and the conditional covariance matrix $\mathbf{C}_R[k]$ is given by

$$\mathbf{C}_R[k] = E\{\bar{\mathbf{R}}[k]\bar{\mathbf{R}}^H[k]|\bar{\mathbf{S}}[k]\} = \bar{\mathbf{S}}_D[k]\mathbf{C}_H\bar{\mathbf{S}}_D^H[k] + N_R\sigma_n^2\mathbf{I}_{TN} \tag{3.10}$$

$\bar{\mathbf{S}}_D[k]$ is a block diagonal matrix

$$\bar{\mathbf{S}}_D[k] = \begin{bmatrix} \mathbf{S}[k+1] & & & \\ & \mathbf{S}[k+2] & & \\ & & \ddots & \\ & & & \mathbf{S}[k+N] \end{bmatrix} \tag{3.11}$$

and $\mathbf{C}_H$ is the covariance matrix of the vector $\bar{\mathbf{H}}$. $\mathbf{C}_H$ can be represented as

$$\mathbf{C}_H = N_R(\mathbf{C}_h \otimes \mathbf{I}_{N_T}) \tag{3.12}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{C}_h$ is given by

$$\mathbf{C}_h = \begin{bmatrix} R_h[0] & R_h[1] & \cdots & R_h[N-1] \\ R_h[-1] & R_h[0] & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ R_h[-N+1] & \cdots & \cdots & R_h[0] \end{bmatrix}. \tag{3.13}$$

In (3.1), when P is large ($P \geq 4$), $\sum_{p=1}^{P} |x_p[n]|^2 \simeq PE_s$ using the large law of numbers. Therefore

$$\mathbf{S}^H[n]\mathbf{S}[n] \doteq TE_s\mathbf{I}_{N_T}. \tag{3.14}$$

For unitary constellations, $\simeq$ is $=$. Since $\det(\mathbf{C}_R[k]) = \det(\mathbf{C}_H\bar{\mathbf{S}}_D^H[k]\bar{\mathbf{S}}_D[k]+N_R\sigma_n^2\mathbf{I}_{N_TN}) \simeq \det(TE_s\mathbf{C}_H\mathbf{I}_{N_TN} + N_R\sigma_n^2\mathbf{I}_{N_TN})$ is almost independent of $\bar{\mathbf{S}}_D[k]$, (3.8) is equivalent to

$$\bar{\mathbf{S}}[k] = \arg\min_{\bar{\mathbf{S}}[k]} \operatorname{tr}\left(\bar{\mathbf{R}}^H[k]\mathbf{C}_R^{-1}[k]\bar{\mathbf{R}}[k]\right). \tag{3.15}$$

Using the identity $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$, (3.15) becomes

$$\begin{aligned} \bar{\mathbf{S}}[k] &= \arg\min_{\bar{\mathbf{S}}[k]} \frac{1}{N_R\sigma_n^2}\operatorname{tr}\left(\bar{\mathbf{R}}^H[k](\mathbf{I}_{TN} - \bar{\mathbf{S}}_D[k](N_R\sigma_n^2\mathbf{C}_H^{-1} + \bar{\mathbf{S}}_D^H[k]\bar{\mathbf{S}}_D[k])^{-1}\bar{\mathbf{S}}_D^H[k])\bar{\mathbf{R}}[k]\right) \\ &= \arg\max_{\bar{\mathbf{S}}[k]} \operatorname{tr}\left(\bar{\mathbf{R}}^H[k]\bar{\mathbf{S}}_D[k](N_R\sigma_n^2\mathbf{C}_H^{-1} + \bar{\mathbf{S}}_D^H[k]\bar{\mathbf{S}}_D[k])^{-1}\bar{\mathbf{S}}_D^H[k]\bar{\mathbf{R}}[k]\right) \\ &= \arg\max_{\bar{\mathbf{S}}[k]} \operatorname{tr}\left(\bar{\mathbf{R}}^H[k]\bar{\mathbf{S}}_D[k]\mathbf{C}\bar{\mathbf{S}}_D^H[k]\bar{\mathbf{R}}[k]\right) \end{aligned} \tag{3.16}$$

where $\mathbf{C} = (N_R\sigma_n^2\mathbf{C}_H^{-1} + \bar{\mathbf{S}}_D^H[k]\bar{\mathbf{S}}_D[k])^{-1}$. Using (3.14), $\mathbf{C} = \mathbf{D}\otimes\mathbf{I}_{N_T}$ via Kronecker product properties, where $\mathbf{D} = N_R\sigma_n^2\mathbf{C}_h^{-1}+TE_s\mathbf{I}_N$ with the (i,j)-th entry $d_{i,j}$. Therefore, (3.16) can be written as

$$\begin{aligned} \bar{\mathbf{S}} &= \arg\max_{\bar{\mathbf{S}}} \operatorname{tr}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} d_{i,j}\mathbf{R}^H[i]\mathbf{S}[i]\mathbf{S}^H[j]\mathbf{R}[j]\right) \\ &= \arg\max_{\bar{\mathbf{S}}} \sum_{i=1}^{N}\sum_{j=1}^{N} d_{i,j}\operatorname{tr}\left(\mathbf{R}^H[i]\mathbf{S}[i]\mathbf{S}^H[j]\mathbf{R}[j]\right) \end{aligned} \tag{3.17}$$

where the second equality comes from the trace property. For brevity, we omit the time index k in (3.17). If P is small so that (3.14) is not valid and x_p is from a non-unitary constellation, the ML decoder is given later.

To further simplify (3.17), we note that $\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A}^H)^H\mathrm{vec}(\mathbf{B})$. For the (i,j)-th term

$$d_{i,j}\mathrm{tr}\left(\mathbf{R}^H[i]\mathbf{S}[i]\mathbf{S}^H[j]\mathbf{R}[j]\right) = d_{i,j}\mathrm{vec}(\mathbf{S}^H[i]\mathbf{R}[i])^H\mathrm{vec}(\mathbf{S}^H[j]\mathbf{R}[j]). \tag{3.18}$$

Substituting (3.3) into $\mathbf{S}^H[j]\mathbf{R}[j]$, we have

$$\begin{aligned}\mathrm{vec}(\mathbf{S}^H[j]\mathbf{R}[j]) =& \sum_{p=1}^{P}\alpha_p[j]\mathrm{vec}\left(\mathbf{A}_p^T\mathbf{R}[j]\right) - \jmath\beta_p[j]\mathrm{vec}\left(\mathbf{B}_p^T\mathbf{R}[j]\right)\\ =&\mathbf{F}_j\mathbf{s}[j]\end{aligned} \tag{3.19}$$

where $\mathbf{s}[j] = [\alpha_1[j],\ldots,\alpha_P[j],\beta_1[j],\ldots,\beta_P[j]]^T$ and $\mathbf{F}_j = [\mathrm{vec}(\mathbf{A}_1^T\mathbf{R}[j]),\ldots,$ $\mathrm{vec}(\mathbf{A}_P^T\mathbf{R}[j]),-\jmath\mathrm{vec}(\mathbf{B}_1^T\mathbf{R}[j]),\ldots,-\jmath\mathrm{vec}(\mathbf{B}_P^T\mathbf{R}[j])]$. Therefore, we can simplify (3.17) as

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}}\mathbf{s}^T\mathbf{G}\mathbf{s} \tag{3.20}$$

where $\mathbf{s} = [\mathbf{s}^T[1],\ldots,\mathbf{s}^T[N]]^T$ and $\mathbf{G}$ is a positive semidefinite block matrix with the (i,j)-th block $[\mathbf{G}]_{i,j} = d_{i,j}\mathbf{F}_i^H\mathbf{F}_j$.

If the coherence time of the channel is larger than NT_B, the channel remains constant during N frames. Using M-PSK constellations, all $d_{i,j}$'s are then equal, and hence (3.17) reduces to the decision metric given in [43]. However, (3.17) is not limited to BPSK as in [43].

Clearly, a phase ambiguity exists in (3.17). Take Alamouti code with M-PSK ($Q_M = \{e^{\jmath 2\pi m/M}\}$, $m = 0,\ldots,M-1$) for example and let

$$\mathbf{\Theta} = \begin{pmatrix} e^{\jmath 2\pi k/M} & 0 \\ 0 & e^{-\jmath 2\pi k/M}\end{pmatrix}, k \in \{0,1,\ldots,M-1\}. \tag{3.21}$$

If $\bar{\mathbf{S}}$ maximizes (3.17), it can be verified that $\bar{\mathbf{S}}(\mathbf{I}_N\otimes\mathbf{\Theta})$ also maximizes (3.17). There are two possible ways to solve such ambiguity. Firstly, only one pilot can be transmitted, i.e., $x_1[1]$. This scheme has high bandwidth efficiency but it has so-called code rotation problem [43]. Only a few OSTBC are identifiable. Hence a pilot block should be transmitted for those rotatable codes, i.e., Alamouti code.

Eq. (3.20) can be solved via SDR [43]. Here, we suggest the use of SD [2]. If $x_p[k]$'s are from unitary constellations, $\mathbf{s}^T\mathbf{s} = PN$. Therefore the maximization problem (3.20) becomes

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \mathbf{s}^T (\eta \mathbf{I}_{2PN} - \mathbf{G})\, \mathbf{s} \tag{3.22}$$

where η is a constant. If η is larger than the maximum eigenvalue of $\mathbf{G}$, $\rho_{\max}$, $\eta\mathbf{I}_{2PN} - \mathbf{G}$ is positive definite. Possible choices of η can be $\rho_{\max} + \sigma_n^2$, $\rho_{\max} + \rho_{\min}$, and $\mathrm{tr}(\mathbf{G})$, where $\rho_{\min}$ is the minimum eigenvalue of $\mathbf{G}$. We use the first choice in the simulation. Let the Cholesky decomposition of $\eta\mathbf{I}_{2PN} - \mathbf{G}$ be $\mathbf{M}$. Eq. (3.22) can then be reduced to

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{M}\mathbf{s}\|^2. \tag{3.23}$$

The quadratic form (3.23) is similar to BLAST type MIMO systems. Therefore it can be solved by using the V-BLAST (Section 1.2.1) or the SD (Section 1.2.2).

Note that for QPSK, each element of $\check{\mathbf{x}}$ is chosen from the set $\{-1, 1\}$. However, for M-PSK ($M > 4$), this does not hold. For any constellations, if $\alpha_p[n]$ is fixed, $\beta_p[n]$ is restricted by the constellation. In SD, when $\alpha_p[n]$ is assigned a value from its candidate set, the candidate set for $\beta_p[n]$ is determined by the bound given by SD and restriction in the constellation. The details of the complex SD is given in Section 5.3.

When $x_p[k]$'s are not from a unitary constellation, let $\xi_{\max}$ and $\xi_{\min}$ be the maximum and minimum modulus of the constellation $\mathcal{Q}$. Eq. (3.16) is equivalent to minimizing

$$g_1(\mathbf{s}) = \eta PN - \mathrm{tr}\left(\bar{\mathbf{R}}^H \bar{\mathbf{S}}_D (N_R \sigma_n^2 \mathbf{C}_H^{-1} + \bar{\mathbf{S}}_D^H \bar{\mathbf{S}}_D)^{-1} \bar{\mathbf{S}}_D^H \bar{\mathbf{R}}\right) \tag{3.24}$$

where $\mathbf{s}$ is defined in (3.20). It can be proven that

$$g_1(\mathbf{s}) \geq \mathbf{s}^T \left(\frac{\eta}{\xi_{\max}^2}\mathbf{I}_{2PN} - \mathbf{G}'\right) \mathbf{s} = g_2(\mathbf{s}). \tag{3.25}$$

where the (i,j)-th block of $\mathbf{G}'$ is $[\mathbf{G}']_{i,j} = d'_{i,j}\mathbf{F}_i^H\mathbf{F}_j$, $d'_{i,j}$ is the (i,j)-th entry of $\mathbf{D}' = N_R\sigma_n^2\mathbf{C}_h^{-1} + T\xi_{\min}^2\mathbf{I}_N$ and $\mathbf{F}_i$ is defined in (3.19). When using SD, we solve

$g_2(\mathbf{s}) \leq g_1(\mathbf{s}) < r^2$. All of the candidates that satisfy $g_2(\mathbf{s}) < r^2$ are found, and the one that makes $g_1(\tilde{\mathbf{x}})$ a minimum is the ML solution. During the search, the bound r^2 can be updated by $g_1(\tilde{\mathbf{s}})$, where $\tilde{\mathbf{s}}$ is a valid candidate within the hyper-sphere.

3.3 Totally Blind Decoders

3.3.1 Different constellations scheme

In [43], pilot symbols are inserted to resolve the phase ambiguity. The non-identifiable code needs a pilot block. These cause a bandwidth loss. In [46], two different PSK-constellations are used to solve the phase ambiguity in blind OFDM detection, which motivates the totally blind decoder in this section.

In the totally blind decoder, two different constellations are used in N consecutive blocks. The two constellations are chosen such that the angles between a point in one constellation and any points in the other constellation are different. QPSK ($\mathcal{Q}_4 = \{e^{\jmath m\pi/2+\pi/4}, m = 0,1,2,3\}$) and 3-PSK ($\mathcal{Q}_3 = \{e^{\jmath 2m\pi/3}, m = 0,1,2\}$) satisfy this property. For example, 3-PSK is used in the $k+1, k+3, \ldots, K+N-1$ blocks and QPSK is used in the remaining blocks. If $\tilde{\mathbf{S}}$ maximizes (3.16) and $\mathbf{S}[k+1]\boldsymbol{\Theta}$ is also feasible, it can be verified that $\mathbf{S}[k+2]\boldsymbol{\Theta}$ is invalid due to the use of different constellations. Specifically, for Alamouti code, define

$$\boldsymbol{\Theta}_1 = \begin{pmatrix} e^{\jmath\frac{2\pi m}{3}} & 0 \\ 0 & e^{-\jmath\frac{2\pi m}{3}} \end{pmatrix}, \boldsymbol{\Theta}_2 = \begin{pmatrix} e^{\jmath\frac{m\pi}{2}+\frac{\pi}{4}} & 0 \\ 0 & e^{-\jmath(\frac{m\pi}{2}+\frac{\pi}{4})} \end{pmatrix}. \tag{3.26}$$

Both $\mathbf{S}[k+1]\boldsymbol{\Theta}_1$ and $\mathbf{S}[k+2]\boldsymbol{\Theta}_2$ are valid. However, there does not exist a $\boldsymbol{\Theta}$ that makes both $\mathbf{S}[k+1]\boldsymbol{\Theta}$ and $\mathbf{S}[k+2]\boldsymbol{\Theta}$ valid. Therefore, (3.16) has a unique solution. QPSK with 5-PSK and 8-PSK with 7-PSK also satisfy the property.

The 3-PSK and QPSK constellations pair is not optimized in [46]. We find the optimal 3-PSK constellation is $\mathcal{Q}_3 = \{1, e^{\jmath 5\pi/8}, e^{-\jmath 5\pi/8}\}$ and the optimal QPSK is $\mathcal{Q}_4 = \{e^{\jmath m\pi/2+\pi/4}, m = 0,1,2,3\}$ by taking into account the phase ambiguity.

The binary bits are mapped to 3-PSK via a punctured convolutional encoder in [46]. Here we introduce a linear block mapping scheme. 3 binary bits are mapped to two 3-PSK symbols, which consists of 9 tuples. The tuple $(0,0)$ is not mapped and therefore it has 0.17 bits loss. When performing ML decoding, this tuple is similar to the parity check bits in linear block code, which can correct the error. Since the gray mapping does not exist for the 3 bits mapping, we develop a quasi-gray mapping scheme. After optimization, we find the suboptimal mapping is given by

$$\begin{aligned}
&100 \rightarrow (1, e^{\jmath\frac{5\pi}{8}}),\ 010 \rightarrow (1, e^{-\jmath\frac{5\pi}{8}}),\ 001 \rightarrow (e^{\jmath\frac{5\pi}{8}}, 1),\\
&000 \rightarrow (e^{\jmath\frac{5\pi}{8}}, e^{\jmath\frac{5\pi}{8}}),\ 011 \rightarrow (e^{\jmath\frac{5\pi}{8}}, e^{-\jmath\frac{5\pi}{8}}),\ 111 \rightarrow (e^{-\jmath\frac{5\pi}{8}}, 1),\\
&101 \rightarrow (e^{-\jmath\frac{5\pi}{8}}, e^{\jmath\frac{5\pi}{8}}),\ 110 \rightarrow (e^{-\jmath\frac{5\pi}{8}}, e^{-\jmath\frac{5\pi}{8}}).
\end{aligned} \tag{3.27}$$

3.3.2 Superimposed pilots scheme

The superposition of pilot and data symbols has been proposed in [42,47]. However, the scoring method for channel estimation in [47] can not achieve the CRB. We use the superimposed pilots to resolve the phase ambiguity. The p-th transmitted symbol in the n-th block can be represented as

$$x_p[n] = \sqrt{\gamma_{n,p}}t_p[n] + \sqrt{1-\gamma_{n,p}}u_p[n] \tag{3.28}$$

where $t_p[n]$ is the known pilot and $u_p[n]$ is a data symbol from $\mathcal{Q}$. The coefficient $\gamma_{n,p}$ denotes the percentage of the power allocated to training. Therefore, the $\mathbf{s}$ in (3.23) can be written as

$$\mathbf{s} = \mathbf{\Gamma}_1\mathbf{t} + \mathbf{\Gamma}_2\mathbf{u} \tag{3.29}$$

where $\mathbf{t}$ and $\mathbf{u}$ are formulated using the real and imaginary parts of $t_p[n]$ and $u_p[n]$ as $\mathbf{s}$, $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are diagonal matrices with diagonal entries $\sqrt{\gamma_{n,p}}$ and $\sqrt{1-\gamma_{n,p}}$, respectively. Eq. (3.23) can then be simplified as

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \|\mathbf{y} - \mathbf{M}'\mathbf{u}\|^2 \tag{3.30}$$

where $\mathbf{y} = -\mathbf{M}\mathbf{\Gamma}_1\mathbf{t}$ and $\mathbf{M}' = \mathbf{M}\mathbf{\Gamma}_2$. It can be readily verified that there is no phase ambiguity in (3.30) due to the non-zero $\mathbf{y}$ without sacrificing the bandwidth efficiency.

3.4 Channel Estimation

After the data symbols have been detected using (3.23) or (3.30), the channel can be MMSE estimated using $\hat{\mathbf{S}}[n]$. If the channel remains constant during N blocks, the MMSE channel estimator is given by

$$\hat{\mathbf{H}} = \left(c \sum_{n=k+1}^{k+N} \sum_{p=2}^{P} |\hat{x}_p[n]|^2 + \frac{\sigma_n^2}{\sigma_h^2} \right)^{-1} \left(\sum_{n=k+1}^{k+N} \mathbf{S}^H[n]\mathbf{R}[n] \right). \tag{3.31}$$

In this case, the CRB of the channel estimate for the joint channel estimation and decoding can be derived as

$$\text{CRB} = \frac{N_T N_R \sigma_n^2 \sigma_h^2 \mathcal{E}}{\sigma_n^2 + \sigma_h^2 \mathcal{E}} \tag{3.32}$$

where $\mathcal{E} = \sum_{n=k+1}^{k+N} \mathcal{E}_n$ is the total power transmitted during N blocks and $\mathcal{E}_n$ is the total power allocated to the n-th block.

When the channel is QS fading, the MMSE channel estimator becomes

$$\hat{\mathbf{H}} = \left(\hat{\bar{\mathbf{S}}}_D^H[n] \hat{\bar{\mathbf{S}}}_D[n] + \sigma_n^2 (\mathbf{C}_h^{-1} \otimes \mathbf{I}_{N_T}) \right)^{-1} \hat{\bar{\mathbf{S}}}_D^H[n] \bar{\mathbf{R}}[n]. \tag{3.33}$$

The corresponding CRB can be obtained as

$$\text{CRB} = N_R \text{tr} \left(\left(\frac{1}{\sigma_n^2} \mathbf{E} \otimes \mathbf{I}_{N_T} + \mathbf{C}_h^{-1} \otimes \mathbf{I}_{N_T} \right)^{-1} \right) = N_R N_T \text{tr} \left(\left(\frac{1}{\sigma_n^2} \mathbf{E} + \mathbf{C}_h^{-1} \right)^{-1} \right) \tag{3.34}$$

where $\mathbf{E} = \text{diag}\{\mathcal{E}_{k+1}, \ldots, \mathcal{E}_{k+N}\}$. A remarkable property of (3.32) and (3.34) is that the CRB does not depend on the power allocated to the pilots or the location of pilots due to the property of OSTBC. However, the training power determines the SNR required to achieve the CRB. The more training power, the lower SNR is needed.

However, from an information point of view, if the channel is almost constant during the N blocks, the optimal power allocation scheme is $\mathcal{E}_{k+1} = \mathcal{E}_{k+2} = \ldots \mathcal{E}_{k+N} = \mathcal{E}/N$ by maximizing the mutual information between the input and the output [47]. But, when the channel fluctuates severely, the optimal power allocation depends on individual channels. To simplify system design, the equal power allocation scheme is suggested.

The estimated channel (3.31) can be used directly to decode the subsequent blocks. Wiener filtering can be applied to (3.33) to predict the channel in the following blocks. In addition, decision-direct algorithm can be used to update the channel. The blind decoder is only needed to stop the error propagation caused by the decision-direct. These will alleviate the computational complexity of the blind decoder.

3.5 OSTBC Decoding over Time-Selective Fading Channels with Perfect CSI

In Sections 3.1-3.4, we have assumed that the channel is quasi-static. However, in practical systems, this assumption may not hold even under normal vehicle speeds. In Section 3.6, we will evaluate our blind decoder in a time-selective channel even though the decoder is derived with quasi-static assumption. We also want to compare with the ML decoder with perfect CSI. But in a time-selective channel, the ML decoding of OSTBC is not simply symbol by symbol decoding.

In the following, we consider a time-selective channel. Since OSTBC with perfect CSI can be decoded in each block or within T time slots, we omit the block index and rewrite (3.5) as

$$r_{t,j} = \sum_{i=1}^{N_T} h_{i,j}[t]s_{t,i} + w_{t,j}. \tag{3.35}$$

Eq. (3.35) can be written in matrix form as

$$\mathbf{r}[t] = \mathbf{H}^T[t]\mathbf{s}[t] + \mathbf{w}[t], \quad ,t = 1, \ldots, T \tag{3.36}$$

where $\mathbf{r}[t] = [r_{t,1}, \ldots, r_{t,N_R}]^T$, $\mathbf{s}[t] = [s_{t,1}, \ldots, s_{t,N_T}]^T$, $\mathbf{H}[t] = [h_{i,j}[t]]$ and $\mathbf{w}[t] = [w_{t,1}, \ldots, w_{t,N_R}]^T$. Using (3.3), we have

$$\mathbf{s}[t] = \sum_{p=1}^{P} \left(\alpha_p \mathbf{a}_{t,p}^T + j\beta_p \mathbf{b}_{t,p}^T\right) = \mathbf{C}[t]\mathbf{s} \tag{3.37}$$

where $\mathbf{a}_{t,p}$ and $\mathbf{b}_{t,p}$ are the t-th row of $\mathbf{A}_p$ and $\mathbf{B}_p$, $\mathbf{s} = [\alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_P]^T$, and

$\mathbf{C}[t] = [\mathbf{a}_{1,p}^T, \ldots, \mathbf{a}_{T,p}^T, j\mathbf{a}_{1,p}^T, \ldots, j\mathbf{a}_{T,p}^T]$. Substituting (3.37) into (3.36), we get

$$\mathbf{r}[t] = \mathbf{H}^T[t]\mathbf{C}[t]\mathbf{s} + \mathbf{w}[t], \quad , t = 1, \ldots, T. \tag{3.38}$$

Assuming the additive noise is white Gaussian, the ML decoder for OSTBC is given by

$$\begin{aligned} \hat{\mathbf{s}} &= \arg\min_{\mathbf{s}} \sum_{t=1}^{T} \|\mathbf{r}[t] - \mathbf{H}^T[t]\mathbf{C}[t]\mathbf{s}\|^2 \\ &= \arg\min_{\mathbf{s}} \|\tilde{\mathbf{r}} - \tilde{\mathbf{H}}\mathbf{s}\|^2 \end{aligned} \tag{3.39}$$

where $\tilde{\mathbf{r}} = [\mathbf{r}^T[1], \ldots, \mathbf{r}^T[T]]^T$ and $\tilde{\mathbf{H}} = [\mathbf{C}[1]^T\mathbf{H}[1], \ldots, \mathbf{C}[T]^T\mathbf{H}[T]]^T$.

Eq. (3.39) is also in a quadratic form similar to BLAST type MIMO systems. It can thus be solved by using the V-BLAST detection algorithm (Section 1.2.1) or the SD (Section 1.2.2).

3.6 Simulation Results

We consider the OSTBC $N_T = 3$ and $P = 4$ [43]

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ -x_2 & x_1 & x_4 & -x_3 \\ -x_3 & -x_4 & x_1 & x_2 \end{pmatrix}. \tag{3.40}$$

The number of receiver antennas is $N_R = 3$ and the number of blocks is $N = 8$. BPSK is used throughout the simulation. $x_1[1]$ is transmitted as a pilot to solve the phase ambiguity. The SNR is defined to be $2E\{\|\mathbf{H}\|_p^2\}/N_0$. ML decoding with perfect CSI using (3.39) is used as the benchmark.

We first consider that the channel remains constant for N blocks. Fig. 3.2 shows the BER versus SNR of various blind decoders, i.e., (3.23) with SD (blind SD), blind SDR [43], blind cyclic [41] and blind subspace [42]. The blind SD and blind SDR perform substantially better than the other blind decoders. At BER=10^{-3}, the blind

SD has a 0.2-dB gain over blind SDR. Compared with the benchmark, the blind SD performs 2 dB worse, which is due to the differential mechanism behind the blind decoder.

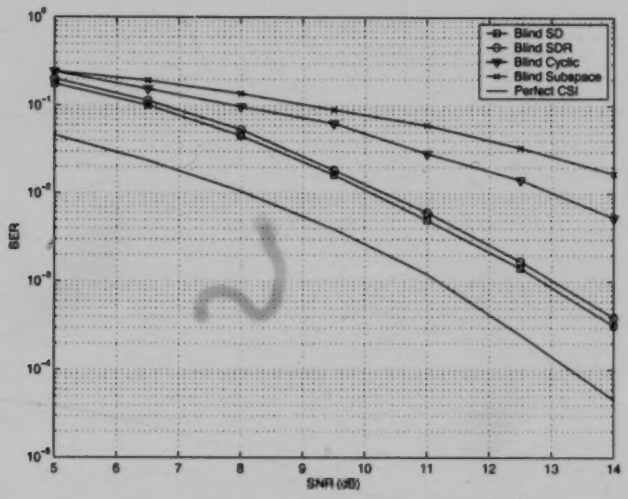

Fig. 3.2. BER versus SNR for different blind decoders with $N = 8$ and BPSK.

Fig. 3.3 compares the average complexities of different blind decoders. The flops is used as the measurement since it is less dependent on the programming skills and the hardware. The complexity of the preprocessing stage such as Cholesky decomposition is also counted for blind SD. The complexities of blind SDR and blind subspace are independent of SNR, while the other two depend on SNR. The blind SDR is the most complex one, though its complexity is claimed to be $\mathcal{O}((NP)^{3.5})$. The reason is that the time constant is very large in blind SDR. In the observed SNR region, the blind SD achieves the smallest complexity. Therefore, blind SD outperforms the other blind decoders in both BER and complexity in that region.

The effects of channel variation on different decoders are shown in Fig. 3.4. The Jakes' model is assumed for the channel and $f_d T_B = 0.005$. The ML SDR denotes solving (3.20) using SDR. The performance of the blind SDR, blind cyclic and blind subspace decoders are greatly degraded due to the unsuitable model used. At BER=10^{-2}, the ML SD still has a 0.2-dB gain over ML SDR. However, the per-

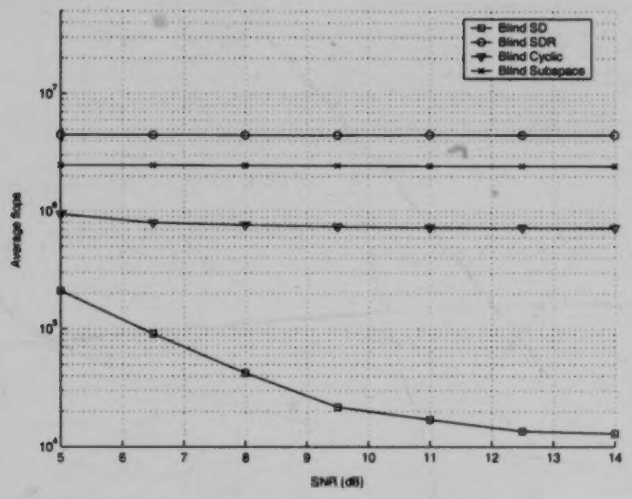

Fig. 3.3. Average flops versus SNR for different blind decoders with $N = 8$ and BPSK.

formance gap between ML SD and the benchmark enlarges to 2.8 dB, which is caused by the QS assumption.

3.7 Conclusion

We have investigated joint channel estimation and decoding for OSTBC without CSI. A general decision metric for a QS channel is derived. Our blind decoder results in a quadratic optimization problem, and it can be efficiently solved using SD, V-BLAST and SDR. Pilot symbols are needed to solve the inherent phase ambiguity. To save the bandwidth, we propose two totally blind decoders using different constellations and superimposed pilots. The MMSE channel estimator and its CRB are given in the end. Power allocation issue is also discussed.

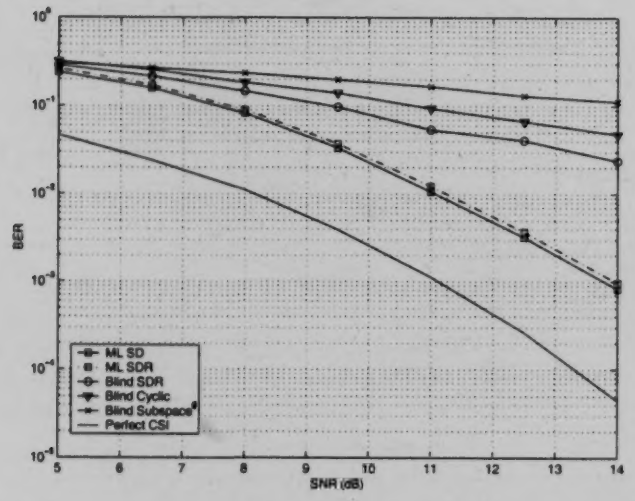

Fig. 3.4. BER versus SNR for different blind decoders with $N = 8$, BPSK and $f_d T_B = 0.005$.

Chapter 4

Efficient Detection of Multiple-Symbol Differential Unitary Space-Time Modulation

This chapter develops a class of optimal, reduced-complexity detectors for differential unitary space-time modulation (DUSTM) called bound intersection detector (BID). This chapter is organized as follows. Section 4.1 introduces the DUSTM. Section 4.2 first develops the BID for single symbol differential detection (SSD) and also presents several algorithms for multiple-symbol differential detection (MSD). In Section 4.3, MSD for DUSTM over Ricean fading channels is derived and discussed. Section 4.4 gives numerical results, and Section 4.5 concludes this chapter.

4.1 Introduction

4.1.1 Background

As can be seen in Chapter 2, the wireless communication system capacity can be substantially enhanced by employing multiple transmit and receive antennas. However,

coherent detection needs perfect channel state information (CSI), which is difficult to obtain in a fast varying mobile environment and/or in a multiple-antenna system, motivating the development of noncoherent detection strategies. Differential space-time modulation (DSTM) has thus received considerable interest [8–10]. Tarokh and Jafarkhani [8] first proposed a DSTM scheme with orthogonal constellations, which can be classified as a nongroup design, existing only for a limited number of antennas. Hochwald [9] developed a general framework for DUSTM, where finite-group properties can simplify the transmitter modulation and constellation design; moreover, diagonal signals, where only one transmit antenna is active at any time, exist for any number of any number of antennas. The reader is referred to [9, 10] for a thorough treatment of DUSTM.

From [9], DUSTM generalizes the classical single-antenna differential phase-shift keying (DPSK), and, similar to DPSK, DUSTM performs 3 dB worse than its coherent counterpart. To improve the performance of single symbol differential detection (SSD), multiple-symbol differential detection (MSD) has been developed for M-ary phase-shift keying (M-PSK) signals transmitted over an additive white Gaussian noise (AWGN) channel [48]. In MSD, $N+1$ consecutive received samples are jointly processed to detect N data symbols. For moderate N, MSD bridges the performance gap between coherent MPSK and MDPSK. The performance of ML-MSD improves with increasing N, albeit at an exponential growth of detection complexity with increasing N. Several low-complexity single-antenna MSD algorithms are developed in [49–51]. Both Mackenthun's algorithm and the improved version [51] only work for AWGN or static fading channels and suffer a mismatch problem [52]. Lampe *et. al.* [52] develop a fast detection algorithm using sphere decoding. Another low-complexity approach, which performs worse than sphere decoding but better than SSD, is decision-feedback differential detection (DF-DD) [53, 54]. These papers treat single-antenna systems only.

Naturally, attempts have been made to extend some of these detection techniques to the multiple-antenna case. In [55], noncoherent DSTM receivers using MSD and

DF-DD are developed to overcome the performance loss in fast fading channels. The robustness of DF-DD to imperfect knowledge of channel parameters is investigated in [56]. However, as the MSD decision rule in [55] is computationally too complex, only the special case of diagonal signals is considered in [55]. A general decision metric for MSD of DUSTM is derived in [57], which uses the Viterbi algorithm for detection, resulting in high complexity for large constellation size L. In both [55] and [57], a major thrust is to analyze the error performance of these schemes, as opposed to developing efficient decoders. The first important paper dealing with this decoding problem is by Clarkson *et al.* [58], which develops a low-complexity approximate algorithm for the SSD of diagonal signals. Their main insight is to recognize that the detection problem can be approximated as a closest vector problem (CVP) in a lattice, as similar problems appear in number theory applications. They use the celebrated LLL lattice reduction algorithm, named after Lenstra, Lenstra, and Lovasz [59]. This, however, results in a suboptimal algorithm; moreover, it cannot be directly applied for MSD. Throughout this chapter, we refer to it as the LLL decoder.

In this chapter, based on [55–57], we derive a decision metric for MSD of DUSTM over a quasi-static (QS) fading channel. The main contribution is, however, a fast exact ML detector, called bound-intersection detector (BID), for single symbol detection with diagonal constellations. Since the decision metric consists of non-negative summands, using a bound, they can be used to generate candidate sets of the transmit signal. The intersection of all such sets constitutes the whole solution space, which is repeatedly pruned until the optimal solution is found. A key novel feature of the BID is the use of the extended Euclidean algorithm [21], well-known for determining the greatest common divisor (gcd) of two integers, to generate the candidate sets. While the ML-search complexity is exponential in the number of transmit antennas and the data rate, our algorithm, particularly in high signal-to-noise ratio (SNR), achieves significant computational savings over the ML algorithm. Interestingly, our BID has lower complexity than the LLL decoder in high SNR (recall that BID is ML while LLL is suboptimal). BID also forms the basic backbone of efficient MSD algorithms; we thus develop four BID variants for MSD. The first two are ML and

use branch-and-bound (BnB), the third one is suboptimal, which first uses BID to generate a candidate subset and then exhaustively searches over the reduced space, and the last one generalizes decision-feedback differential detection.

We then generalize the optimal decision rule to MIMO Ricean channels. To significantly reduce this detection complexity, a suboptimal MSD-based DF-DD is proposed using our BID. Although DF-DD does not achieve ML performance, it achieves substantial performance improvement over the conventional differential detector (CDD) while its complexity is only linear in N. Furthermore, we combine the BnB principle and BID and give a so-called sphere decoding bound intersection detector (SD-BID) to efficiently solve the MSD problem, which offers ML performance. In the high SNR region, the complexity of SD-BID is even lower than that of the DF-DD.

4.1.2 System model and differential unitary space-time modulation

We consider a MIMO system with N_T transmit and N_R receive antennas, and the input-output relationship can be written as [9,10,55,57]

$$\mathbf{R}[n] = \mathbf{S}[n]\mathbf{H}[n] + \mathbf{W}[n] \tag{4.1}$$

where $\mathbf{S}[n] = [s_{i,j}[n]]$ is the $T \times N_T$ transmitted matrix during the n-th interval, and T is the number of time slots per block interval. $s_{i,j}[n]$ $i = 1, 2, \ldots, T$ and $j = 1, 2, \ldots, N_T$ is transmitted by the j-th antenna in the i-th slot. $\mathbf{R}[n] = [r_{i,j}[n]]$ is the $T \times N_R$ received signal during the n-th block interval, and $r_{i,j}[n]$ is defined similarly to $s_{i,j}[n]$. $\mathbf{H}[n] = [h_{i,j}[n]]$ is the $N_T \times N_R$ MIMO channel matrix during the n-th block interval. $\mathbf{W}[n] = [w_{i,j}[n]]$ is the $T \times N_R$ noise matrix with independent and identically distributed entries $w_{i,j} \sim \mathcal{CN}(0, \sigma_n^2)$ where σ_n^2 is adjusted to ensure a given average SNR. Each time slot occupies an interval T_s, and the block interval is $T_B = T_s T$, both in seconds.

For a frequency-flat Rayleigh fading MIMO channel and a rich scattering envi-

ronment [60], the entries $h_{i,j}[n] \sim \mathcal{CN}(0,1)$ for $i = 1,\ldots,N_T$ and $j = 1,\ldots,T$. The autocovariance of the channel gains is given by

$$E\left\{h_{i,j}[n]h^*_{i',j'}[n+m]\right\} = \delta_{i,i'}\delta_{j,j'}R_h[m] \tag{4.2}$$

where $\delta_{i,j}$ is the Kronecker delta and $R_h[m]$ is the correlation function of $h_{i,j}[n]$. This model describes spatially independent and identically distributed random channel gains with the identical correlation function $R_h[m]$. The fading channel is QS, i.e., the underlying continuous fading channel gain $h_{i,j}(t)$ remains constant over each block interval, and hence $h_{i,j}[n]$ is approximated by the mid-point sample of $h_{i,j}(t)$ [57], whereas the channel changes from block to block. Typically, when Clarke's (Jake's) model [45] is used, $R_h[m] = J_0(2\pi m f_d T_B)$, where $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind given by

$$J_0(x) = \sum_{k=0}^{\infty}(-1)^k \frac{x^{2k}}{4^k(k!)^2}, \tag{4.3}$$

and f_d is the Doppler spread due to users' mobility. Note that the QS condition is met when $f_d T_s < 0.03$ [55].

For a frequency-flat Ricean fading MIMO channel, $h_{i,j}[n]$ can be expressed as the summation of the direct component $(h_d)_{i,j}[n]$ and the scattered component $(h_s)_{i,j}[n]$

$$h_{i,j}[n] = (h_d)_{i,j}[n] + (h_s)_{i,j}[n]. \tag{4.4}$$

Assuming the Rice factor K is common to all paths, K is defined as $|(h_d)_{i,j}|^2/E\{|(h_s)_{i,j}|^2\}$ [61]. The autocovariance of the channel gains can be obtained similarly as (4.2). Eq. (4.2) can be written as

$$\mathbf{R}[n] = \mathbf{S}[n](\mathbf{H}_d[n] + \mathbf{H}_s[n]) + \mathbf{W}[n] \tag{4.5}$$

where $[\mathbf{H}_d]_{i,j}[n] = (h_d)_{i,j}[n]$ and $[\mathbf{H}_s]_{i,j}[n] = (h_s)_{i,j}[n]$.

From [9], the transmit symbols $\mathbf{S}[n]$ are generated using a finite group $\mathcal{V} = \{\mathbf{V}_l, l = 0,1,\ldots,L-1\}$, where $\mathbf{V}_l$ is a $T \times N_T$ unitary matrix ($\mathbf{V}_l\mathbf{V}_l^H = \mathbf{I}_T$), and $L = 2^{N_T R}$, where R denotes the data rate. We assume $T = N_T$, and $\mathbf{V}_0 = \mathbf{I}_{N_T}$. $N_T R$ binary

information bits are first converted to an integer $l \in [0, L-1]$, and $\mathbf{V}[n] = \mathbf{V}_l$ is chosen from $\mathcal{V}$. The n-th transmitted block is encoded as

$$\mathbf{S}[n] = \mathbf{V}[n]\mathbf{S}[n-1]. \tag{4.6}$$

In the first block, $\mathbf{S}[0] = \mathbf{V}_0$ is sent. The internal composition property of a group ensures that $\mathbf{S}[n] \in \mathcal{V}$, and is unitary for any positive n. Specifically, for diagonal constellations, the unitary matrices $\mathbf{V}_l$ are chosen as

$$\mathbf{V}_l = \text{diag}\left\{e^{j2\pi u_1 l/L}, e^{j2\pi u_2 l/L}, \ldots, e^{j2\pi u_{N_T} l/L}\right\} \tag{4.7}$$

where u_i, $i = 1, \ldots, N_T$ are optimized to achieve the maximum diversity product [9].

4.2 Reduced Complexity Differential Unitary Space-Time Demodulation over Rayleigh MIMO Channels

4.2.1 Decision metric

We first derive the ML MSD metric for DUSTM over Rayleigh MIMO channels. Since MSD estimates the transmitted symbols in N consecutive intervals given $N+1$ received symbols, let us consider symbol intervals $n = k$ to $n = N + k$. Let $\bar{\mathbf{R}}[k] = [\mathbf{R}^H[k], \mathbf{R}^H[k+1], \ldots, \mathbf{R}^H[k+N]]^H$ and $\bar{\mathbf{V}}[k] = [\mathbf{V}^H[k+1], \mathbf{V}^H[k+2], \ldots, \mathbf{V}^H[k+N]]^H$. The ML estimate for $\bar{\mathbf{V}}[k]$ can be expressed as

$$\hat{\bar{\mathbf{V}}}[k] = \arg\max_{\bar{\mathbf{V}}[k]} f(\bar{\mathbf{R}}[k] | \bar{\mathbf{V}}[k]) \tag{4.8}$$

where $f(a|b)$ is the probability density function (pdf) of a conditioned on b. We define the covariance matrix of $[h_{i,j}[k], \ldots, h_{i,j}[k+N]]^T$ as

$$\mathbf{C}_h = \begin{bmatrix} R_h[0] & R_h[1] & \cdots & R_h[N] \\ R_h[-1] & R_h[0] & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ R_h[-N] & \ldots & \ldots & R_h[0] \end{bmatrix} \tag{4.9}$$

and let $\mathbf{A} = (\mathbf{C}_h + \sigma_n^2 \mathbf{I}_{N+1})^{-1}$. Using the results of [55–57], (4.8) can be simplified as

$$\widehat{\mathbf{V}}[k] = \arg\max_{\tilde{\mathbf{V}}[k]} \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} a_{i,j} \mathrm{Re} \left\{ \mathrm{tr} \left(\mathbf{R}^H[i+k-1] \left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right)^H \right) \mathbf{R}[j+k-1] \right\} \tag{4.10}$$

where $a_{i,j}$ is the (i,j)-th entry of $\mathbf{A}$. We normalize $a_{i,j}$ with $a_m = \max_k |a_{k,k+1}|$, $k = 1, \ldots, N$ or $a_m = a_{\lfloor \frac{N}{2} \rfloor, \lfloor \frac{N}{2} \rfloor + 1}$ and denote $\tilde{a}_{i,j} = -a_{i,j}/a_m$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x (the reason for this normalization will be clear soon). Eq. (4.10) is equivalent to

$$\begin{aligned} \widehat{\mathbf{V}}[k] &= \arg\min_{\tilde{\mathbf{V}}[k]} \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} -\tilde{a}_{i,j} \mathrm{Re} \left\{ \mathrm{tr} \left(\mathbf{R}^H[i+k-1] \left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right)^H \mathbf{R}[j+k-1] \right) \right\} \\ &= \arg\min_{\tilde{\mathbf{V}}[k]} \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} \left\| \mathbf{R}[j+k-1] - \tilde{a}_{i,j} \left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right) \mathbf{R}[i+k-1] \right\|_F^2 . \end{aligned} \tag{4.11}$$

When the channel is static over the $N+1$ blocks or equivalently $R_h[n] = 1$, it can be readily obtained that $\tilde{a}_{i,j} = 1$ ($i = 1, 2, \ldots, N$, $j = 2, \ldots, N+1$ and $i \neq j$). Eq. (4.11) becomes

$$\widehat{\mathbf{V}}[k] = \arg\min_{\tilde{\mathbf{V}}[k]} \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} \left\| \mathbf{R}[j+k-1] - \left(\prod_{m=i+k}^{j+k-1} \mathbf{V}[m] \right) \mathbf{R}[i+k-1] \right\|_F^2 . \tag{4.12}$$

When $N = 1$, (4.11) reduces to

$$\widehat{\mathbf{V}}[k+1] = \arg\min_{\tilde{\mathbf{V}}[k+1]} \|\mathbf{R}[k+1] - \mathbf{V}[k+1]\mathbf{R}[k]\|_F^2 . \tag{4.13}$$

Eq. (4.13) is the decision rule given in [9, Eq. (21)]. Hence, the differential detector in [9] is still ML in a QS fading channel. If the normalization in (4.11) is not performed as in [55, 57], the decision rule will not reduce to [9, Eq. (21)] when $N = 1$. The normalization will provide a tighter bound, as will be shown in the next section. Eq. (4.12) can be interpreted as the summation of ML metrics between any two received symbols within the $N+1$ receive blocks. The non-negative summands in (4.11) facilitate our efficient MSD algorithm in Section 4.2.2.

If the channel changes in each time slot, the MSD metric is derived in [55] for diagonal constellations. It can be readily verified that (4.10) reduces to (26) in [55]. Hence, the efficient detection algorithms in Section 4.2.2 can also be applied to the noncoherent receivers in [55].

4.2.2 Reduced complexity single-symbol detection

To put the development of our new algorithm in perspective, let us briefly review the problem and several previous contributions. The key idea of [58] is to convert the decoding of diagonal differential constellations to the CVP in a modular lattice via the cosine approximation ($\cos\alpha \approx 1-\alpha^2/2$). An n-dimensional lattice L generated by a set of linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^k$ is the set $L = \{\sum a_i \mathbf{v}_i \mid a_i \in \mathbb{Z}\}$. Given a lattice L and arbitrary vector $\mathbf{y}$, the CVP is to find $\mathbf{x} \in L$ so that $\|\mathbf{x}-\mathbf{y}\|_p$ is the minimum where the distance is measured in l_p norm $1 \leq p < \infty$. The shortest vector problem (SVP) is the homogeneous version of the CVP (i.e., $\mathbf{y}$ is the origin). Both these problems are known to be NP-hard. Recent results show the CVP in an n-dimensional lattice to be NP-hard to approximate to within factor $n^{c/\log\log n}$ for some constant $c > 0$ [62]. Note also that [58, Eq. (12)] involves translation from a modular lattice to a non-modular lattice. Such a translation has also been considered in [63] and [64]. The celebrated LLL algorithm [59] is a polynomial-time algorithm that approximates both the SVP and CVP to within a factor of $2^{O(n)}$. Thus the LLL decoder [58] is faster than the ML exhaustive search, but the cosine approximation and the LLL algorithm incur a performance loss. The CVP can also be optimally solved by the well-known SD [13]. In [65], SD has thus been used along with the lattice approximation of [58]. But note that the search space increases to L^{N_T} in [65] while the original search space is only L. In addition, the cosine approximation also makes the SD solution suboptimal. Therefore, the direct application of SD is not optimal in terms of both computational complexity and performance. We next derive a novel, efficient SSD algorithm by combining bounding and the extended Euclidean algorithm.

The ML SSD rule (4.13) for diagonal signals can be written as

$$\hat{l} = \arg\min_{l} \left\| \mathbf{R}[k+1] - \mathbf{V}_l^l \mathbf{R}[k] \right\|_F^2 \tag{4.14}$$

where $\mathbf{V}_1$ is defined in (4.7). The cost metric in (4.14) can be expanded as

$$\begin{aligned}
\hat{l} &= \arg\min_{l} \sum_{i=1}^{N_T} \sum_{j=1}^{N_R} \left| r_{i,j}[k+1] - e^{j2\pi u_i l/L} r_{i,j}[k] \right|^2 \\
&= \arg\min_{l} \sum_{i=1}^{N_T} \sum_{j=1}^{N_R} |r_{i,j}[k+1]|^2 + |r_{i,j}[k]|^2 - 2\mathrm{Re}\left\{ r_{i,j}^*[k+1] r_{i,j}[k] e^{j2\pi u_i l/L} \right\} \\
&= \arg\min_{l} \sum_{i=1}^{N_T} A_i - B_i \cos\left[(u_i l - \phi_i) 2\pi/L \right] = \arg\min_{l} \varphi(l)
\end{aligned} \tag{4.15}$$

where

$$\begin{aligned}
A_i &= \sum_{j=1}^{N_R} |r_{i,j}[k+1]|^2 + |r_{i,j}[k]|^2 \\
B_i &= 2 \left| \sum_{j=1}^{N_R} r_{i,j}^*[k+1] r_{i,j}[k] \right| \\
\phi_i &= \arg\left(\sum_{j=1}^{N_R} r_{i,j}[k+1] r_{i,j}^*[k] \right) L/2\pi.
\end{aligned} \tag{4.16}$$

We let the arg operation take values in $[0, 2\pi)$ so that $\phi_i \in [0, L)$. If l is the true solution, the cost metric (4.14) becomes

$$e = \sum_{i=1}^{N_T} \sum_{j=1}^{N_R} \left| w_{i,j}[k+1] - e^{j2\pi u_i l/L} w_{i,j}[k] \right|^2 \tag{4.17}$$

where $w_{i,j}[n]$ are the AWGN terms in (4.1). Note that e/σ_n^2 is a chi-square random variable with $2N_T N_R$ degrees of freedom. Therefore, we can choose a bound C to be proportional to the variance of the noise as

$$C = \alpha \sigma_n^2 \tag{4.18}$$

so that the probability that at least one candidate $\hat{l}$ exists, which ensures that the cost metric (4.14) is less than C, is very high:

$$\int_0^{\alpha} \frac{x^{N_T N_R - 1} e^{-x/2}}{\Gamma(N_T N_R) 2^{N_T N_R}} \mathrm{d}x = 1 - \epsilon \tag{4.19}$$

where ϵ is set to a value close to 0 (e.g., $\epsilon = 0.1$). Instead of searching all of the $0 \leq l < L$, we only search the values of l such that $\varphi(l) < C$. To find all of the l's that meet this condition, we note that $\varphi(l)$ (4.15) consists of N_T non-negative terms. Thus a necessary condition for $\varphi(l) < C$ is that each term of (4.15) is less than C, or equivalently

$$A_i - B_i \cos\left[(\text{mod}(u_i l, L) - \phi_i) 2\pi / L\right] < C,\ i = 1, 2, \ldots, N_T. \tag{4.20}$$

where mod (x, L) reduces x to an integer between 0 and L. Let us define the candidate set $\mathcal{L}_i = \{l \,|\, A_i - B_i \cos\left[(\text{mod}(u_i l, L) - \phi_i) 2\pi / L\right] < C,\ 0 \leq l < L\}$, i.e., $\mathcal{L}_i$ consists of all l which satisfy the i-th term in (4.20). Note that when $(A_i - C)/B_i > 1$, $\mathcal{L}_i$ is a null set. If $(A_i - C)/B_i \leq -1$, all the integers in $[0, L)$ are included in $\mathcal{L}_i$. The problem at hand is to determine $\mathcal{L}_i$ efficiently for all i. Since $\cos\theta$ is monotonically decreasing between 0 to π and monotonically increases from π to 2π, and since $\cos\theta$ is an even function, we can readily show that (4.20) is equivalent to

$$|\text{mod}(u_i l, L) - \phi_i| < \rho_i \quad \text{or} \quad L - \rho_i < |\text{mod}(u_i l, L) - \phi_i| < L,\ i = 1, 2, \ldots, N_T \tag{4.21}$$

where

$$\rho_i = \frac{L}{2\pi} \cos^{-1}\left(\frac{A_i - C}{B_i}\right), \tag{4.22}$$

and $\cos^{-1}(x)$ takes values in $[0, \pi]$. Expanding (4.21), we have

$$-\rho_i + \phi_i < \text{mod}(u_i l, L) < \rho_i + \phi_i \tag{4.23a}$$

$$\text{or} \quad L - \rho_i + \phi_i < \text{mod}(u_i l, L) < L + \phi_i \tag{4.23b}$$

$$\text{or} \quad \phi_i - L < \text{mod}(u_i l, L) < \rho_i + \phi_i - L. \tag{4.23c}$$

Define $\mathcal{S}_a = \{l \,|\, -\rho_i + \phi_i < \text{mod}(u_i l, L) < \rho_i + \phi_i\}$, $\mathcal{S}_b = \{l \,|\, L - \rho_i + \phi_i < \text{mod}(u_i l, L) < L + \phi_i\}$ and $\mathcal{S}_c = \{l \,|\, \phi_i - L < \text{mod}(u_i l, L) < \rho_i + \phi_i - L\}$. Since $0 \leq \text{mod}(u_i l, L) < L$ and $0 \leq \phi_i < L$, we have $\mathcal{S}_a = \{l \,|\, \max(-\rho_i + \phi_i, 0) < \text{mod}(u_i l, L) < \min(\rho_i + \phi_i, L)\}$, $\mathcal{S}_b = \{l \,|\, L - \rho_i + \phi_i < \text{mod}(u_i l, L) < L\}$ and $\mathcal{S}_c = \{l \,|\, 0 < \text{mod}(u_i l, L) < \rho_i + \phi_i - L\}$. Clearly, $\mathcal{L}_i = \mathcal{S}_a \cup \mathcal{S}_b \cup \mathcal{S}_c$, and exhaustive checking of (4.23) for all $0 \leq l < L$ is not efficient. We next develop an efficient algorithm to determine $\mathcal{L}_i$.

To do so, we first show how to find l such that $\mathrm{mod}(u_i l, L) = 1$. Since u_i is relatively prime to L, the gcd of integers u_i and L is 1. The well-known *Extended Euclidean Algorithm* [21] computes the gcd of u_i and L, as well as the numbers d_i and k such that

$$u_i d_i + kL = 1 \tag{4.24}$$

where 1 is the gcd of u_i and L. For the details of the Extended Euclidean Algorithm, the reader is referred to [21]. To find $\mathrm{mod}(u_i l, L) = n$, we multiply both sides of (4.24) by n, which yields

$$u_i(nd_i) + knL = n. \tag{4.25}$$

Therefore $l = \mathrm{mod}(nd_i, L)$ satisfies $\mathrm{mod}(u_i l, L) = n$. We are now in a position to determine $\mathcal{L}_i$. Define

$$UB_i = \lfloor \phi_i + \rho_i \rfloor, \quad LB_i = \lceil \phi_i - \rho_i \rceil \tag{4.26}$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x, and $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. We now find that

$$\begin{aligned}
\mathcal{S}_a &= \{\mathrm{mod}(nd_i, L) \mid \max(LB_i, 0) \le n \le \min(UB_i, L)\} \\
\mathcal{S}_b &= \{\mathrm{mod}(nd_i, L) \mid L + LB_i \le n \le L\} \\
\mathcal{S}_c &= \{\mathrm{mod}(n_i d_i, L) \mid 0 \le n \le UB_i - L\}.
\end{aligned} \tag{4.27}$$

Note that it can be readily verified that

$$\mathrm{mod}(nd_i, L) = \mathrm{mod}((n + L)d_i, L) = \mathrm{mod}((n - L)d_i, L). \tag{4.28}$$

Eq. (4.27) immediately reduces to

$$\begin{aligned}
\mathcal{S}_a &= \{\mathrm{mod}(nd_i, L) \mid \max(LB_i, 0) \le n \le \min(UB_i, L)\} \\
\mathcal{S}_b = \{\mathrm{mod}(nd_i, L) \mid LB_i \le n \le 0\}, &\quad \mathcal{S}_c = \{\mathrm{mod}(n_i d_i, L) \mid L \le n \le UB_i\}.
\end{aligned} \tag{4.29}$$

Therefore, we have

$$\mathcal{L}_i = \mathcal{S}_a \bigcup \mathcal{S}_b \bigcup \mathcal{S}_c = \{\mathrm{mod}(n_i d_i, L) \mid LB_i \le n \le UB_i\} \tag{4.30}$$

Let $\mathbf{b}_i = [LB_i, LB_i + 1, \ldots, UB_i]$. The candidate set of l for the i-th term of (4.15) is given by

$$\mathcal{L}_i = \mathrm{mod}(d_i \mathbf{b}_i, L),\ i = 1, 2, \ldots, N_T. \tag{4.31}$$

Although the extended Euclidean algorithm is NP-complete [66], d_i can be computed before detection. Furthermore, (4.31) is a one-to-one mapping. For $\mathbf{B} = [0, 1, \ldots, L-1]$, we can store $\mathbf{M} = \text{mod}(d_i\mathbf{B}, L)$ in memory and (4.31) can be accomplished by $\mathcal{L}_i = [\mathbf{M}(LB_i), \mathbf{M}(LB_i + 1), \ldots, \mathbf{M}(UB_i)]$.

The candidates that satisfy all of the N_T equations (4.20) are chosen, i.e., the candidate set is the intersection of all of the N_T sets $\mathcal{L}_i$ as

$$\mathcal{L} = \bigcap_{i=1}^{N_T} \mathcal{L}_i. \tag{4.32}$$

Intuitively, the term in (4.20) with the largest u_i, which typically is u_{N_T}, varies most with the change of l. Thus, the element in $\mathcal{L}$ that is closest to ϕ_{N_T} is searched first. If no l can make $\varphi(l)$ (4.15) less than the bound C, or equivalently if $\mathcal{L}$ is a null set, we increase the probability $1 - \epsilon$ (e.g., $\epsilon = 0.1^2, 0.1^3, \ldots$), adjust the bound C and perform the same process again. If l^* is chosen, C is replaced by the new cost $\varphi(l^*)$, and l^* is deleted from the set $\mathcal{L}$ ($\mathcal{L} = \mathcal{L} - \{l^*\}$). All $\mathcal{L}_i$ $i = 1, \ldots, N_T$ are updated using the new bound C. In later iterations, (4.32) is replaced by

$$\mathcal{L} = \mathcal{L} \cap \left(\bigcap_{i=1}^{N_T} \mathcal{L}_i \right), \tag{4.33}$$

which avoids duplicate searches and reduces the search space. The process continues until $\mathcal{L}$ becomes the null set. The l with the minimum cost is then the optimal solution. We call this optimal detection algorithm Bound-Intersection Detector (BID).

To further improve the BID performance, we note that each term of (4.15) has a lower bound as

$$lb_i = A_i - B_i \cos\left[\Delta\phi_i 2\pi / L\right], \; i = 1, 2, \ldots, N_T \tag{4.34}$$

where $\Delta\phi_i = \phi_i - \lceil \phi_i \rfloor$ and $\lceil \cdot \rfloor$ denotes the nearest integer to its argument. Hence, the lower and upper bounds (4.26) are updated to

$$\begin{aligned} UB_i &= \left\lfloor \phi_i + \frac{L}{2\pi} \cos^{-1} \left(\frac{A_i - C + \sum_{j=1, j\neq i}^{N_T} lb_j}{B_i} \right) \right\rfloor \\ LB_i &= \left\lceil \phi_i - \frac{L}{2\pi} \cos^{-1} \left(\frac{A_i - C + \sum_{j=1, j\neq i}^{N_T} lb_j}{B_i} \right) \right\rceil. \end{aligned} \tag{4.35}$$

Remarks:

- In high SNR, C is small. The $\mathcal{L}$ in (4.33) usually contains only one element. On the contrary, C becomes large in low SNR. The size of $\mathcal{L}$ approaches L. Therefore, similar to SD, the complexity of BID decreases with the increase of SNR.

- If the normalization in (4.11) is formulated, e/σ_n^2 in (4.17) is not a chi-square random variable with $N_T N_R$ degrees of freedom. The initial bound C will be loose and difficult to estimate.

- The dimension of the lattice formed in [58] is the product $N_T N_R$. Hence, for a large number of receive antennas N_R, the complexity of LLL may indeed be larger than that of a brute-force ML search which is linear in $L = 2^{RN_T}$. However, our problem formulation (4.15) does not expand the search space.

- In [65], SD has been used to solve the DUSTM detection problem based on the lattice formulation in [58]. But note that the search space increases to L^{N_T} in the lattice representation while the original search space is only L.

- The bottleneck of BID is the computation of N_T candidate sets $\mathcal{L}_i$, and since they can be obtained simultaneously, the BID algorithm can be readily parallelized, an attractive feature for the implementation on a float-point multiple-processor digital signal processor (DSP).

The pseudo code of the BID is given in **Algorithm 3**.

4.2.3 Reduced complexity multiple-symbol detection

For the MSD of the diagonal signals, the search space increases to L^N, and the computation of the metric (4.11) is more complex than in the SSD case. Eq. (4.11)

input : Received signals $\mathbf{R}[k+1]$, $\mathbf{R}[k]$.
output: The optimal l.

Compute A_i, B_i and ϕ_i; $\epsilon = 0.1$;
Compute α using (4.29) and $C = \alpha\sigma_n^2$;
for $i \leftarrow 1$ *to* N_T **do**
    if $(A_i - C)/B_i > 1$ **then**
        $\mathcal{L} = \phi$; **goto** 16;
    else if $(A_i - C)/B_i \leq -1$ **then**
        $\mathcal{L} = (0 : L-1)$;
    else
        Compute LB_i and UB_i using (4.26) and $\mathcal{L}_i$ using (4.31);
    end
end
$\mathcal{L} = \bigcap_{i=1}^{N_T} \mathcal{L}_i$;
if $\mathcal{L} == \phi$ **then**
    $\epsilon = 0.1\epsilon$; **goto** 2;
end
Sort $\mathcal{L}$ according to $|\text{mod}(u_{N_T} l, L) - \phi_{N_T}|$; $l^* = \mathcal{L}(1)$; $l_{min} = l^*$;
$C = \|\mathbf{R}[k+1] - \mathbf{V}_1^{l^*}\mathbf{R}[k]\|_F^2$; $\mathcal{L} = \mathcal{L} - \{l^*\}$;
while $\mathcal{L} \neq \phi$ **do**
    for $i \leftarrow 1$ *to* N_T **do**
        if $(A_i - C)/B_i > 1$ **then**
            $\mathcal{L} = \phi$; **goto** 33;
        else if $(A_i - C)/B_i \leq -1$ **then**
            $\mathcal{L}_i = (0 : L-1)$;
        else
            Compute LB_i and UB_i using (4.26) and $\mathcal{L}_i$ using (4.31);
        end
    end
    $\mathcal{L} = \mathcal{L} \cap \bigcap_{i=1}^{N_T} \mathcal{L}_i$;
    if $\mathcal{L} == \phi$ **then**
        return l_{min};
    end
    $l^* = \mathcal{L}(1)$; $\mathcal{L} = \mathcal{L} - \{l^*\}$;
    if $\|\mathbf{R}[k+1] - \mathbf{V}_1^{l^*}\mathbf{R}[k]\|_F^2 < C$ **then**
        $C = \|\mathbf{R}[k+1] - \mathbf{V}_1^{l^*}\mathbf{R}[k]\|_F^2$; $l_{min} = l^*$;
    else
        goto 33;
    end
end
return l_{min};

Algorithm 3: Bound Intersection Detection Algorithm

can be reduced to

$$\begin{aligned}\hat{\mathbf{l}} &= [\hat{l}_{k+1}, \hat{l}_{k+2}, \ldots, \hat{l}_{k+N}] \\ &= \underset{l_{k+1}, l_{k+2}, \ldots, l_{k+N}}{\arg\min} \sum_{i=1}^{N} \sum_{j=i+1}^{N+1} \left\| \mathbf{R}[j+k-1] - \tilde{a}_{i,j} \mathbf{V}_1^{(\sum_{m=i+k}^{j+k-1} l_m)} \mathbf{R}[i+k-1] \right\|_F^2 .\end{aligned} \tag{4.36}$$

where $\mathbf{V}_1$ is given in (4.7). We next give four MSD algorithms which generalize the BID algorithm.

4.2.3.1 MSD1

We first use BID for the SSD of the N block symbols, and the result is denoted by $\hat{\mathbf{l}}$. $\hat{\mathbf{l}}$ is then substituted back into (4.36), and the cost is denoted by C. Note that (4.36) is the summation of non-negative terms. The exhaustive search is performed. After each of the $(N+1)N/2$ terms in (4.36) is computed, the current cost is compared with C. If it is larger than C, the search stops, and another candidate is tested. When all the $(N+1)N/2$ terms have been finished, the total cost is compared with C. If the cost is less than C, C is replaced by this value, the current $\mathbf{l}$ is saved, and the search continues until all of the L^N possible candidates have been finished. The best one is output as the optimal solution. This MSD is similar to a BnB algorithm. Unfortunately, this algorithm is not very efficient when L is large and when the SNR is low, making the initial bound C loose.

4.2.3.2 MSD2

The efficiency of our proposed BID is due to avoiding search of all $0 \leq l < L$. To apply the same idea to MSD, we also begin by using BID for applying SSD for N symbols over $N+1$ blocks. The resulting $\hat{\mathbf{l}}$ is then substituted back into (4.36), and the initial bound is obtained as C. Since (4.36) is the summation of $(N+1)N/2$ non-negative terms, a necessary condition for the cost (4.36) be less than C is that

each term of (4.15) is less than C, in particular

$$\left\|\mathbf{R}[k+N] - \tilde{a}_{N-1,N}\mathbf{V}_1^{l_{k+N}}\mathbf{R}[k+N-1]\right\|_F^2 < C, \tag{4.37}$$

which is the SSD problem. The candidate set $\mathcal{L}_{k+1}$ for l_{k+N} can be found by using BID. For every l_{k+1} in $\mathcal{L}_{k+N}$, the bound for l_{k+N-1} can be improved to $C - B_{k+N}$, where $B_{k+N} = \left\|\mathbf{R}[k+N] - \mathbf{V}_1^{l_{k+N}}\mathbf{R}[k+N-1]\right\|_F^2$. The candidate set $\mathcal{L}_{k+N-1}$ for l_{k+N-1} can also be found by using BID. The similar process continues for l_{k+N-2}, and so on. When it comes to $l_{m'}$, the bound is updated as $C - \sum_{i=m'+1}^{k+N} B_i$, where B_i is given by

$$B_i = \sum_{j=i+1}^{N+1} \left\|\mathbf{R}[j+k-1] - \tilde{a}_{i,j}\mathbf{V}_1^{(\sum_{m=i+1}^{j} l_m)}\mathbf{R}[i+k-1]\right\|_F^2. \tag{4.38}$$

When a set of $\mathbf{l}$ has been chosen, (4.36) is computed and compared with C. If it is less than C, C is updated, $\mathbf{l}$ is saved and deleted from their candidate set, and the candidate set for each l_m is updated by using the new bound. The process continues until all of the elements in the candidate set have been searched. The output is the optimal solution.

However, the initial bound can become loose with the increase of N and in high SNR. In this case, the candidate set usually contains all of the l's. To overcome this problem and further reduce the complexity, the idea similar to BID can be used to find the lower bound of each term in (4.36), which can be obtained the same as in (4.34). The bound for each l can be further improved by using these lower bounds (details omitted for brevity).

4.2.3.3 MSD3

We may use the output of SSD as a starting point of MSD. In [49], a reduced-complexity detector is proposed for MSD of M-PSK. The key idea is to search for a small candidate subset with the $s > 1$ largest symbol-wise metrics for pairs of received

signals, and then search exhaustively over the reduced space of size s^N where N is the number of M-PSK symbols. Similarly, for the MSD of DUSTM, we first modify BID to do SSD for each signal l_m $(m = k+1, \ldots, k+N)$ and generate a candidate list of the $s > 1$ smallest metrics – instead of returning only the optimal solution. This can be accomplished by choosing a larger initial bound. If less than the s candidates are found, the bound is increased until they are obtained. The s^N N-tuples are substituted into (4.36), and the one with minimum cost is output. When s is small, the number of N-tuples to search is relatively small, and this significantly reduces complexity. Furthermore, when testing all of the s^N N-tuples, the BnB algorithm in MSD1 can also be used to further reduce the complexity. Reference [49] shows that when choosing $s = 2$, the performance of the reduced-complexity algorithm is nearly ML. While ignoring the first stage BID, the complexity of the reduced-complexity MSD is only $(s/L)^N$ of that of the ML search. The effectiveness of MSD3 in static fading channels is verified in Section 4.4.

4.2.3.4 MSD4

The MSD problem (4.11) can be formulated as maximum likelihood sequence estimation (MLSE). DF-DD [55] is equivalent to a decision feedback sequence estimator (DFSE). In [67] and [68], a reduced-state sequence estimator (RSSE) is introduced to reduce the number of states in MLSE. DFSE can also be viewed as a special RSSE. Similarly, a reduced-state differential detector (RS-DD) can be used to solve (4.11) as a generalization of DF-DD.

As a special case of DFSE in [67,68], RS-DD replaces $l_{k+1}, \ldots, l_{k+M}$ with previously decided symbols $\hat{l}_{k+1}, \ldots, \hat{l}_{k+M}$, $0 \leq M \leq N-1$. The ML detection is then performed for $l_{k+M+1}, \ldots, l_{k+N}$. Clearly, if $M = N-1$, RS-DD reduces to DF-DD in [55] and if $M = 0$, RS-DD reduces to MSD. For $l_{k+M+1}, \ldots, l_{k+N}$, MSD1 and MSD2 can be used to reduce the complexity of exhaustive search. Therefore, RS-DD or MSD4 gives a tradeoff between performance and complexity.

4.3 Reduced Complexity Differential Unitary Space-Time Demodulation over Ricean MIMO Channels

4.3.1 Decision metric

Similarly, we consider the sequence from $n = k$ to $n = k + N$. Let $\bar{\mathbf{R}}[k] = [\mathbf{R}^H[k], \mathbf{R}^H[k+1], \ldots, \mathbf{R}^H[k+N]]^H$ and $\bar{\mathbf{H}}[k] = [\mathbf{H}^H[k], \mathbf{H}^H[k+1], \ldots, \mathbf{H}^H[k+N]]^H$. The input-output relationship for the N symbols can be expressed as

$$\begin{aligned}\bar{\mathbf{R}}[k] =& \bar{\mathbf{S}}_D[k]\bar{\mathbf{H}}[k] + \bar{\mathbf{W}}[k] \\ =& \bar{\mathbf{S}}_D[k](\bar{\mathbf{H}}_d[k] + \bar{\mathbf{H}}_s[k]) + \bar{\mathbf{W}}[k]\end{aligned} \tag{4.39}$$

where $\bar{\mathbf{S}}_D[k]$ is a block diagonal matrix

$$\bar{\mathbf{S}}_D[k] = \begin{bmatrix} \mathbf{S}[k] & & & \\ & \mathbf{S}[k+1] & & \\ & & \ddots & \\ & & & \mathbf{S}[k+N] \end{bmatrix} \tag{4.40}$$

and $\bar{\mathbf{H}}_d[k] = [\mathbf{H}_d^H[k], \mathbf{H}_d^H[k+1], \ldots, \mathbf{H}_d^H[k+N]]^H$, $\bar{\mathbf{H}}_s[k] = [\mathbf{H}_s^H[k], \mathbf{H}_s^H[k+1], \ldots, \mathbf{H}_s^H[k+N]]^H$, $\bar{\mathbf{W}}[k] = [\mathbf{W}^H[k], \mathbf{W}^H[k+1], \ldots, \mathbf{W}^H[k+N]]^H$. $\mathrm{vec}(\bar{\mathbf{R}}[k])$ is a complex Gaussian vector and the conditional pdf given $\bar{\mathbf{S}}_D[k]$ is

$$f(\bar{\mathbf{R}}|\,\bar{\mathbf{S}}_D) = \frac{1}{(\pi^{NN_T}\det(\mathbf{C}_R))^{N_R}} \exp\left\{-\mathrm{tr}\left((\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d)^H\mathbf{C}_R^{-1}(\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d)\right)\right\}. \tag{4.41}$$

We ignore the time index k in (4.41) for simplicity. The autocovariance matrix $\mathbf{C}_R$ is given by

$$\mathbf{C}_R = E\{\bar{\mathbf{R}}\bar{\mathbf{R}}^H\} = \bar{\mathbf{S}}_D\mathbf{C}_H\bar{\mathbf{S}}_D^H + N_R\sigma_n^2\mathbf{I}_{N_TN} \tag{4.42}$$

where $\mathbf{C}_H$ is the covariance matrix of $\bar{\mathbf{H}}$ and can be represented as

$$\mathbf{C}_H = N_R(\mathbf{C}_h \otimes \mathbf{I}_{N_T}) \tag{4.43}$$

where $\otimes$ denotes the Kronecker product [69] and $\mathbf{C}_h$ is given in (4.9). Since $\mathbf{S}[n]$'s $(n = k, k+1, \ldots, k+N)$ are unitary matrix, $\bar{\mathbf{S}}_D \bar{\mathbf{S}}_D^H = \mathbf{I}_{N_T N}$. We have

$$\begin{aligned}
\mathbf{C}_R =& \bar{\mathbf{S}}_D \mathbf{C}_H \bar{\mathbf{S}}_D^H + N_R \sigma_n^2 \mathbf{I}_{N_T N} \\
=& N_R \bar{\mathbf{S}}_D (\mathbf{C}_h \otimes \mathbf{I}_{N_T} + \sigma_n^2 \mathbf{I}_{N_T N}) \bar{\mathbf{S}}_D^H \\
=& N_R \bar{\mathbf{S}}_D \left[(\mathbf{C}_h + \sigma_n^2 \mathbf{I}_N) \otimes \mathbf{I}_{N_T} \right] \bar{\mathbf{S}}_D^H \\
=& N_R \bar{\mathbf{S}}_D \left(\mathbf{C} \otimes \mathbf{I}_{N_T} \right) \bar{\mathbf{S}}_D^H
\end{aligned} \tag{4.44}$$

where $\mathbf{C} = \mathbf{C}_h + \sigma_n^2 \mathbf{I}_{N+1}$. The third line (4.44) follows the distributivity property of Kronecker product. It can be readily verified that $\det(\mathbf{C}_R)$ does not depend on $\bar{\mathbf{S}}_D$. Therefore maximizing (4.41) is equivalent to minimizing

$$g(\bar{\mathbf{S}}_D) = \mathrm{tr}\left((\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d)^H \mathbf{C}_R^{-1} (\bar{\mathbf{R}} - \bar{\mathbf{S}}_D \bar{\mathbf{H}}_d) \right). \tag{4.45}$$

Note that

$$\begin{aligned}
\mathbf{C}_R^{-1} =& \frac{1}{N_R} \bar{\mathbf{S}}_D \left(\mathbf{C} \otimes \mathbf{I}_{N_T} \right)^{-1} \bar{\mathbf{S}}_D^H \\
=& \frac{1}{N_R} \bar{\mathbf{S}}_D \left(\mathbf{C}^{-1} \otimes \mathbf{I}_{N_T} \right) \bar{\mathbf{S}}_D^H.
\end{aligned} \tag{4.46}$$

The first equality comes from $\bar{\mathbf{S}}_D \bar{\mathbf{S}}_D^H = \mathbf{I}_{N_T N}$ $(\bar{\mathbf{S}}_D^H \bar{\mathbf{S}}_D = \mathbf{I}_{N_T N})$ and the second equality comes from the Kronecker product property $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ ($\mathbf{A}$ and $\mathbf{B}$ are square nonsingular matrices) [69]. We Cholesky factorize $\mathbf{C}^{-1}$ as $\mathbf{C}^{-1} = \mathbf{U}^H \mathbf{U}$, $\mathbf{U}$ is upper triangular. Using the Kronecker product property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, $\mathbf{C}^{-1} \otimes \mathbf{I}_{N_T} = (\mathbf{U} \otimes \mathbf{I}_{N_T})^H (\mathbf{U} \otimes \mathbf{I}_{N_T}) = \bar{\mathbf{U}}^H \bar{\mathbf{U}}$ and $\bar{\mathbf{U}}$ is also upper triangular. This factorization needs to be done only once. After some manipulations and ignoring constants, (4.45) can be simplified as

$$g(\bar{\mathbf{S}}_D) = \left\| \bar{\mathbf{U}} \bar{\mathbf{H}}_d - \bar{\mathbf{U}} \bar{\mathbf{S}}_D^H \bar{\mathbf{R}} \right\|_F^2 = \left\| \bar{\mathbf{Y}} - \bar{\mathbf{U}} \bar{\mathbf{S}}_D^H \bar{\mathbf{R}} \right\|_F^2 \tag{4.47}$$

where $\bar{\mathbf{Y}} = \bar{\mathbf{U}} \bar{\mathbf{H}}_d = [\mathbf{Y}^H[k], \mathbf{Y}^H[k+1], \ldots, \mathbf{Y}^H[k+N]]^H$ and $\mathbf{Y}[n]$ is an $N_T \times N_R$ matrix. The MSD for DUSTM over MIMO Ricean channels is given by

$$\left\{ \hat{\mathbf{S}}[k], \ldots, \hat{\mathbf{S}}[k+N] \right\} = \underset{\mathbf{S}[k], \ldots, \mathbf{S}[k+N] \in \mathcal{V}}{\arg\min} \left\| \bar{\mathbf{Y}} - \bar{\mathbf{U}} \bar{\mathbf{S}}_D^H \bar{\mathbf{R}} \right\|_F^2. \tag{4.48}$$

The transmitted signals can be differentially detected as

$$\hat{\mathbf{V}}[n] = \hat{\mathbf{S}}[n+1] \hat{\mathbf{S}}^*[n] \tag{4.49}$$

When $N_T = N_R = 1$, the MSD (4.48) for MIMO systems reduces to the MSD for single input and single output (SISO) systems.

Remarks:

- When $K = 0 \Rightarrow \mathbf{H}_d = \mathbf{0}$, (4.48) reduces to the decision metric in 4.11, corresponding to the Rayleigh fading channels. When $K \to \infty \Rightarrow \sigma_h^2 \to 0$, (4.48) is equivalent to

$$\left\{\hat{\mathbf{S}}[k], \ldots, \hat{\mathbf{S}}[k+N]\right\} = \underset{\mathbf{S}[k],\ldots,\mathbf{S}[k+N]\in\mathcal{V}}{\arg\min} \|\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d\|^2. \tag{4.50}$$

  Eq. (4.50) corresponds to the coherent detection with perfect CSI. Similarly, with different Rice factor K, the MSD decision metric (4.48) varies between Rayleigh fading and perfect CSI cases.

- If $N = 2$, we find that (4.48) reduces to the CDD [9] as

$$\hat{\mathbf{V}}[k] = \arg\min_{\mathbf{V}[k]\in\mathcal{V}} \|\mathbf{R}[k] - \mathbf{V}[k]\mathbf{R}[k]\|_F^2. \tag{4.51}$$

- If $K \neq 0$, the decision metric (4.48) is variant to a phase shift common to all the components in $\mathbf{S}_D$. Therefore different from the Rayleigh case, the exhaustive search need to test all the L^N candidate vectors instead of L^{N-1}.

- Increasing K the MSD performs more like a coherent detector, which is linear in N. But solving (4.50) needs exhaustively search over L. We need to use BID to reduce the complexity for (4.50).

4.3.2 Reduced complexity multiple-symbol detection

We now present our sphere decoding bound intersection detector (SD-BID) to solve the MSD (4.48). Like SD, we only examine the candidates that satisfy

$$\|\bar{\mathbf{Y}} - \bar{\mathbf{U}}\bar{\mathbf{S}}_D\bar{\mathbf{R}}\|^2 \leq R^2. \tag{4.52}$$

Let the entries of $\mathbf{U}$ be denoted by $u_{i,j}$, $i \leq j$. Taking the upper triangular and Kronecker product structure of $\bar{\mathbf{U}}$ into account, (4.52) can be written as

$$\sum_{i=1}^{N} \left\| \mathbf{Y}[k+i] - \sum_{j=i}^{N} u_{i,j} \mathbf{S}[k+j] \mathbf{R}[k+j] \right\|_F^2 \leq R^2 \tag{4.53}$$

Thus a necessary condition for (4.53) is

$$\|\mathbf{Y}[k+N] - u_{N,N} \mathbf{S}[k+N] \mathbf{R}[k+N]\|_F^2 \leq R^2 \tag{4.54}$$

$$\sum_{i=N-1}^{N} \left\| \mathbf{Y}[k+i] - \sum_{j=i}^{N} u_{i,j} \mathbf{S}[k+j] \mathbf{R}[k+j] \right\|_F^2 \leq R^2 \tag{4.55}$$

$$\vdots \qquad \vdots$$

$$\sum_{i=1}^{N} \left\| \mathbf{Y}[k+i] - \sum_{j=i}^{N} u_{i,j} \mathbf{S}[k+j] \mathbf{R}[k+j] \right\|_F^2 \leq R^2. \tag{4.56}$$

Conditions (4.54)-(4.56) can be checked componentwise. To proceed, we start from $\mathbf{S}[k+N]$. Using BID, we can obtain its candidate set

$$\mathcal{I}_N = \left\{ \mathbf{V}_1^l \,\middle|\, \|\mathbf{Y}[k+N] - u_{N,N} \mathbf{V}_1^l \mathbf{R}[k+N]\|_F^2 \leq R^2, l \in \{0, 1, \ldots, L-1\} \right\} \tag{4.57}$$

where $\mathbf{V}_1$ is defined in (4.7). If $\hat{\mathbf{S}}[k+N]$ is chosen from $\mathcal{I}_N$, it is substituted into (4.55). The candidate set for s_{N-1} is

$$\mathcal{I}_{N-1} = \left\{ \mathbf{V}_1^l \,\middle|\, \left\| \mathbf{Y}[k+N-1] - u_{N-1,N-1} \mathbf{V}_1^l \mathbf{R}[k+N-1] - u_{N-1,N} \hat{\mathbf{S}}[k+N] \mathbf{R}[k+N] \right\|_F^2 \leq R^2, \right.$$
$$\left. l \in \{0, 1, \ldots, L-1\} \right\} \tag{4.58}$$

After choosing $\hat{\mathbf{S}}[k+j]$ for $\mathbf{S}[k+j]$ from their candidate set, $i+1 \leq j \leq N$. we define

$$d_m^2 = \left\| \mathbf{Y}[k+m] - \sum_{j=m}^{N} u_{i,j} \mathbf{S}[k+j] \mathbf{R}[k+j] \right\|_F^2 \tag{4.59}$$

$$R_N^2 = R^2, \; R_m^2 = R_{m+1}^2 - d_{m+1}^2, \; i \leq m \leq N-1 \tag{4.60}$$

The candidate set for $\mathbf{S}[k+i]$ can be obtained as

$$\mathcal{I}_i = \left\{ \mathbf{V}_1^l \left| \left\| \mathbf{Y}[k+i] - u_{i,i}\mathbf{V}_1^l \mathbf{R}[k+i] - \sum_{j=k}^{N} u_{k,j}\hat{\mathbf{S}}[k+j]\mathbf{R}[k+j] \right\|^2 \le R_k^2, \right. \right.$$
$$\left. \left. l \in \{0, 1, \ldots, L-1\} \right\} \right. . \tag{4.61}$$

When all $\hat{\mathbf{S}}[k+i]$ has found, all the R_i's are updated according to

$$R_N^2 = \|\bar{\mathbf{Y}} - \bar{\mathbf{U}}\bar{\mathbf{S}}_D\bar{\mathbf{R}}\|^2,\ R_i^2 = R_{i+1}^2 - d_{i+1}^2,\ i = N-1, \ldots, 1. \tag{4.62}$$

The same process continues until all the candidates meet (4.52) have been checked. The best candidate is output as the ML solution. Eqs. (4.52)-(4.62) are identical to the corresponding operations in SD. The only differences are in SD-BID we use Frobenius norm and the candidate set is obtained using BID. If $N_T = N_R = 1$, the SD-BID reduces to the SD.

The initial radius R can also be obtained according to the statistic of $g(\bar{\mathbf{S}}_D)$ in (4.45)

$$g(\bar{\mathbf{S}}_D) = \mathrm{tr}\left((\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d)^H \mathbf{C}_R^{-1}(\bar{\mathbf{R}} - \bar{\mathbf{S}}_D\bar{\mathbf{H}}_d)\right). \tag{4.63}$$

If $\bar{\mathbf{S}}_D$ is the true solution, using (4.5), $\mathbf{X} = \bar{\mathbf{S}}_D^H[k]\bar{\mathbf{R}}[k] - \bar{\mathbf{H}}_d[k] = \bar{\mathbf{H}}_s[k] + \mathbf{S}_D^H[k]\bar{\mathbf{W}}[k]$ is zero mean complex Gaussian with autocovariance matrix $\mathbf{C}_X = \mathbf{C}_h + \sigma_n^2\mathbf{I}_N$. Therefore $e = \mathrm{tr}\{\mathbf{X}^H(\mathbf{C}_h + \sigma_n^2\mathbf{I}_N)^{-1}\mathbf{X}\}$ is a chi-square random variable with $2NN_RN_T$ degrees of freedom. As in (4.20), R^2 can be chosen to make the probability that e is less than R^2 very high.

$$\int_0^{R^2} \frac{x^{NN_RN_T-1}e^{-x/2}}{\Gamma(NN_RN_T)2^{NN_RN_T}}\mathrm{d}x = 1 - \epsilon. \tag{4.64}$$

ϵ can be reduced to enlarge R^2 to make sure that the ML solution can be found. The initial radius here does not depend on the noise variance. Hence, we do not need to estimate the noise variance.

The Schnorr and Euchner (SE) strategy [14] can also be generalized to SD-BID. In each step, we not only find the candidate set $\mathcal{I}_i$ but also compute the corresponding d_i^2 using (4.59) and store it in $\mathcal{D}_i$. $\mathcal{I}_i$ is sorted according to $\mathcal{D}_i$. The candidate with minimum d_i^2 is searched first.

Assuming correct decisions of $\hat{\mathbf{S}}[k], \ldots, \hat{\mathbf{S}}[k+N-1]$, the MSD for DUSTM (4.48) can be readily modified to MSD based DF-DD by replacing $\mathbf{S}[k], \ldots, \mathbf{S}[k+N-1]$ in (4.48) with $\hat{\mathbf{S}}[k], \ldots, \hat{\mathbf{S}}[k+N-1]$. Our BID can be used to solve the DF-DD. We also note that decision feedback sequence estimator is a special case of the RSSE [67,68]. Similarly, a reduced-state differential detector (RS-DD) can be used to solve (4.48) as a generalization of the DF-DD. Instead of assuming $N-1$ correct feedbacks in (4.48), RS-DD only uses M $(0 \leq M \leq N-1)$ decision feedbacks. $\mathbf{S}[k], \ldots, \mathbf{S}[k+M]$ in (4.48) are replaced with $\hat{\mathbf{S}}[k], \ldots, \hat{\mathbf{S}}[k+M]$ and SD-BID is used for the $N-M$ dimensions problem. If $M=0$, the RS-DD reduces to SD-BID and DF-DD when $M=N-1$. Thus, both the performance and complexity of RS-DD are between SD-BID and DF-DD.

4.4 Simulation Results

We now provide and discuss simulation results. We assume a MIMO channel model as described in Section 4.1.2 and generate the channel gains by sampling a continuous fading process via the Jakes' model [45]. We use the diagonal signals with parameters u_i, $i=1,\ldots,N_T$ from [58, Table 1]. The brute-force ML detector is referred to as ML detector in the following. For coherent detection (CD), the transmit symbols $\mathbf{S}[n]$ are estimated assuming perfect knowledge of the channel matrix $\mathbf{H}[n]$. The information symbols are recovered by differential decoding.

4.4.1 Rayleigh channels

We first show the results over Rayleigh fading channels. Fig. 4.1 compares the performance of BID for SSD with that of ML and the LLL decoder [58] when $N_T = 3,4,5$, $N_R = 1$ and $R=2$. Our proposed BID performs exactly ML. At a BER of 10^{-3} and $N_T = 3$, the LLL decoder performs 0.15 dB worse than the ML decoder.

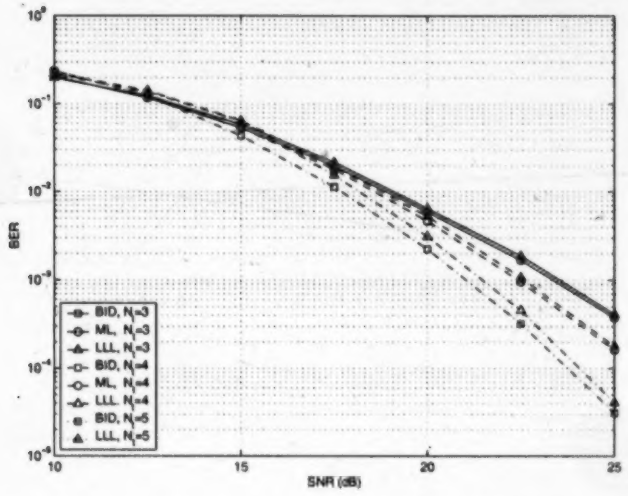

Fig. 4.1. Performance comparison for $N_T = 3, 4, 5$ transmitter antennas, $N_R = 1$ receiver antenna as a function of SNR. The channel is static fading and $R = 2$.

When $N_T = 5$, the performance loss by using LLL decoder increases to 0.5 dB at BER=10^{-3}. As stated in [59], LLL achieves an approximation factor $2^{O(n)}$, which is exponential in the dimension n, which agrees with our simulation results that the gap between ML and LLL increases with the increase of N_T. Note that in [58] an exact decoder is also proposed. This exact algorithm may be used together with LLL, which increases the complexity by a factor of $2^{(N_T+1)N_T/4+N_T}$. Moreover, the exact decoder incurs a performance loss due to cosine approximation.

Fig. 4.2 shows the complexity of BID in flops when $N_T = 2, 3, 4, 5$, $N_R = 1$ and $R = 2$. We use the flops function (it provides an estimate of the number of floating point operations performed by a sequence of Matlab statements) in Matlab to compare the numerical efficiency of various decoders. We do not consider parallelization issues. The LLL decoder follows exactly the one given in [58] without using their exact algorithm. With the increasing SNR, the flops of BID reduce significantly (Fig. 4.2). The ML and LLL complexities are almost constant, given in Table 4.1 for comparison. In high SNR, our proposed BID is much more efficient than both ML and LLL while offering ML performance. The flops required by the ML decoder is between 10 to

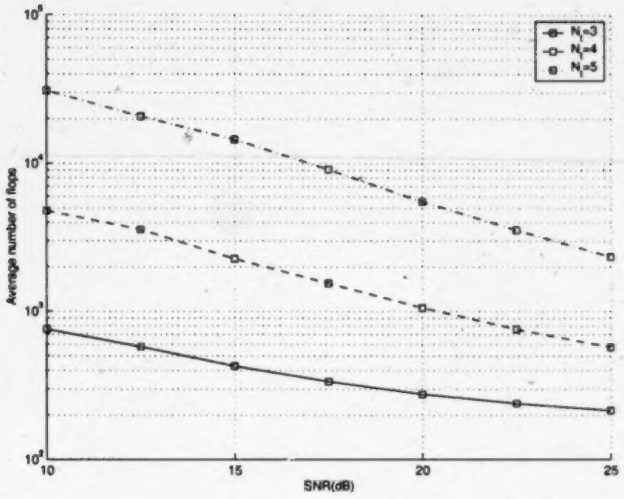

Fig. 4.2. Complexity of BID for $N_T = 3, 4, 5$ transmitter antennas, $N_R = 1$ receiver antenna versus SNR. The channel is static fading and $R = 2$.

30 times of that of BID. With the increase of N_T, the complexity gap between our BID and LLL decreases while the performance gap increases. Note that DUSTM is especially effective in this region [9], which is consistent with the more efficient efficient region of BID, making it especially suitable for DUSTM.

Fig. 4.3 compares the complexity of BID and LLL with a fixed number of transmit antennas $N_T = 4$, a different number of received antennas N_R and $R = 2$. In [58], the lattice dimension increases as $N_T N_R$. For large N_R, the complexity of LLL on

TABLE 4.1

Complexity comparison for ML, LLL and BID in flops.

	ML	LLL	BID (25dB)
N_T=3, N_R=1, R=2	2469	720	214
N_T=4, N_R=1, R=2	13312	2561	572
N_T=5, N_R=1, R=2	66560	8462	2341
N_T=6, N_R=1, R=2	344064	19864	14765

the naive lattice formulation is much more than that of BID. However, we compare the LLL for our formulation (4.15) which only has N_T terms regardless of N_R. The complexity of LLL is almost independent of N_R, and the complexity difference is due to the preprocessing step. Interestingly, in low SNR, the complexity of BID decreases for large N_R, while the complexity of BID increases in high SNR. Each term in (4.15) is a combination of N_R terms. In low SNR, a larger N_R requires a larger value for each term in (4.15), resulting in a smaller candidate set. Since we count the preprocessing flops in computing A_i, B_i and ϕ_i in (4.16), larger N_R encounters higher complexity to compute these parameters and the complexity is dominant by the preprocessing step. In fact, the performance relates to the complexity in our BID. The probability of finding the true solution reflects the tightness of the bound.

Fig. 4.4 illustrates the performance improvement of MSD for a static fading channel. A MIMO system with $N_T = 4$, $N_R = 1$ and $R = 1$ is simulated. The performance gap between $N = 6$ and $N \to \infty$ is relatively small [55]. Hence, in our simulation $N = 3, 6$ blocks of received signals are collected for detection. Since both MSD1 and MSD2 are ML, the performance of MSD2 only is shown in Fig. 4.4. The performance of MSD2 is compared with those of SSD, CD, DF-DD, MSD3 and MSD4. In MSD3 and MSD4, we choose $s = 2$ and $M = 3$, respectively. The performance loss over MSD2 when using MSD3 is negligible even when $s = 2$, which verifies the effectiveness of MSD3. The gap between DF-DD and MSD2 is also small. At BER= 10^{-5}, the gap is only 0.2 dB for $N = 3, 6$. When $N = 3$ blocks are used, MSD2 has a 1-dB performance gain over SSD at BER= 10^{-5}, and when $N = 6$ blocks are used, the performance gain increases to 1.8 dB. MSD2 with $N = 6$ has only a 0.8-dB loss over CD. Fig. 4.5 compares the complexity of different detectors in a static fading channel in terms of the average flops per block. In high SNR, the complexity of MSD1, MSD2, MSD3 and MSD4 decreases, a common property of BnB detectors since their performance depends on the noise variance or equivalently the SNR. In high SNR, the complexity of MSD2 is the lowest among all the detectors. The complexity of MSD1 is high since it only performs naive BnB. The high complexity of DF-DD is due to the computation of canceling the previous symbols, but DF-DD

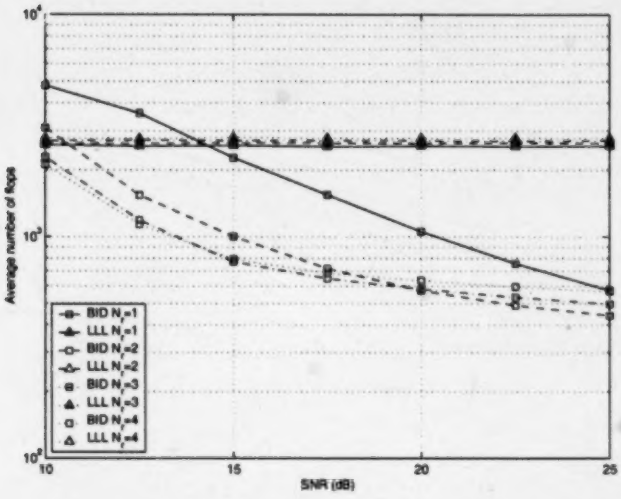

Fig. 4.3. Complexity comparison between BID and LLL for $N_T = 4$ transmitter antennas, $N_R = 1, 2, 3, 4$ receiver antenna versus SNR. The channel is static fading and $R = 2$.

cannot offer ML performance. Both MSD3 and MSD4 have lower complexity than MSD1 and MSD2 in low SNR, and MSD3 and MSD4 perform better than DF-DD. Therefore, MSD3 and MSD4 are suitable in low SNR, and MSD2 is efficient in high SNR.

In Figs. 4.6 and 4.7, the performance and complexity are compared for different detectors with $f_d T_s = 0.0075$, and the other parameters are set the same as in Figs. 4.4 and 4.5. An error floor appears for SSD in high SNR. When $N = 3$, the performance gap between MSD2 and DF-DD is 0.4 dB at BER= 10^{-6}, and the gap increases to 1 dB when $N = 6$ (Fig. 4.6). However, both MSD3 and MSD4 perform close to MSD2. MSD2 has a 6.5-dB loss over CD with $N = 3$ at BER= 10^{-6}, but the loss reduces to 2.5 dB when $N = 6$. The complexity of different detectors as shown in Fig. 4.7 has similar properties as those explained in Fig. 4.6.

We compare the performance and complexity for different detectors in Figs. 4.8 and 4.9 with $f_d T_s = 0.03$, and the other parameters are the same as before. Note that

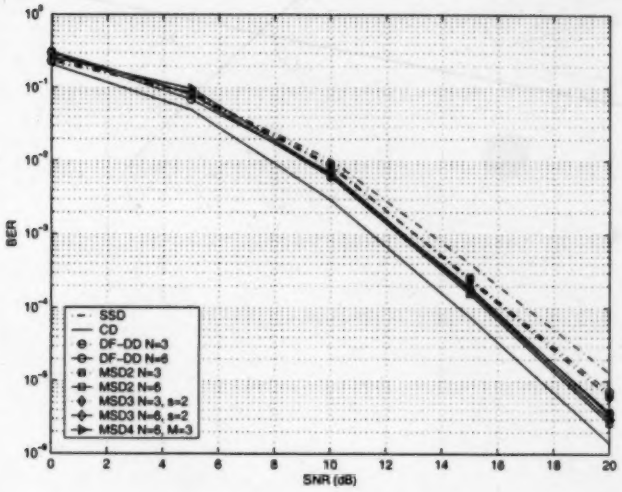

Fig. 4.4. Performance comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ and $R = 1$ as a function of SNR. The channel is constant within N blocks.

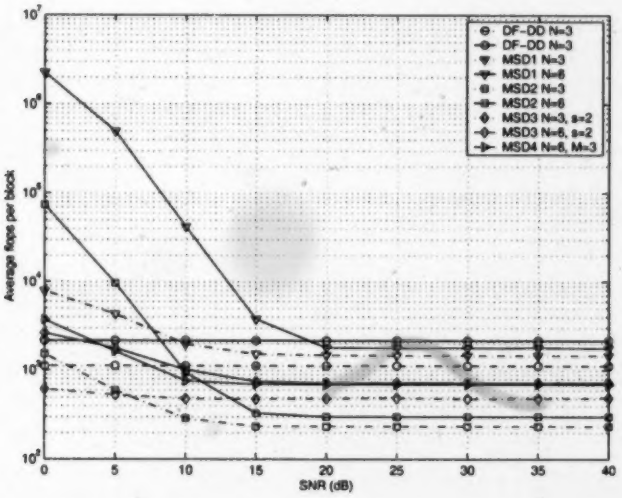

Fig. 4.5. Complexity comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ and $R = 1$ as a function of SNR. The channel is constant within N blocks.

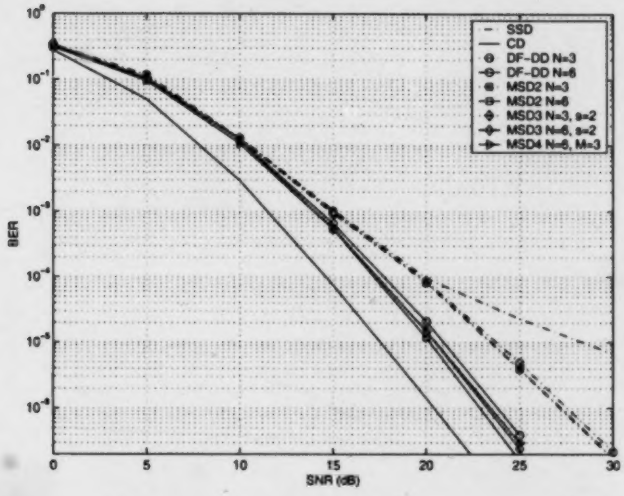

Fig. 4.6. Performance comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ and $R = 1$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.0075$ and $R = 1$.

SSD exhibits a large error floor, which can be reduced by using both DF-DD and our proposed MSD's. MSD2 has smaller error floors than DF-DD. When $N = 6$, the error floor is not observed for both MSD2 and MSD4 within the plotted SNR region. There also exists a large gap between MSD2 and DF-DD. MSD4 performs close to MSD2; for example, at BER= 10^{-7}, the performance gap is only 1.2 dB. MSD3 also exhibits large error floors, which can be reduced by increasing both s and N. Compared to the case $f_d T_s = 0.0075$, the performance gap between MSD2 with $N = 6$ and CD increases significantly: almost 10 dB at BER= 10^{-6}. Therefore, the performance of all of the noncoherent detectors degrades with increasing $f_d T_s$. The complexity of all of the proposed MSD's increases with increasing $f_d T_s$. This is because for large $f_d T_s$ the coefficients $\tilde{a}_{i,j}$ in (4.11) are far from 1. The bound given by computing (4.11) will be on average larger than that in small $f_d T_s$. MSD2 still achieves the minimum complexity in high SNR. In low SNR, MSD3 and MSD4 again have lower complexity than MSD2. They are suitable in low SNR where the complexity of MSD2 is rather high.

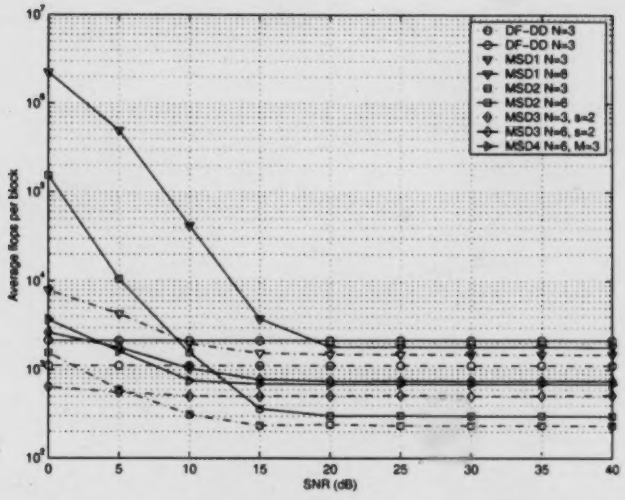

Fig. 4.7. Complexity comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.0075$ and $R = 1$.

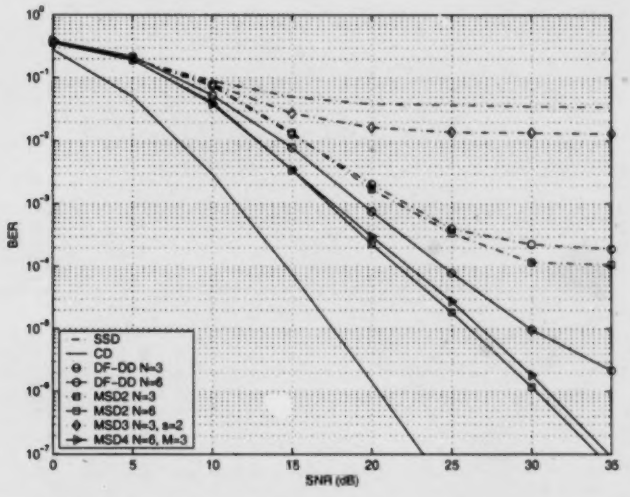

Fig. 4.8. Performance comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.03$ and $R = 1$.

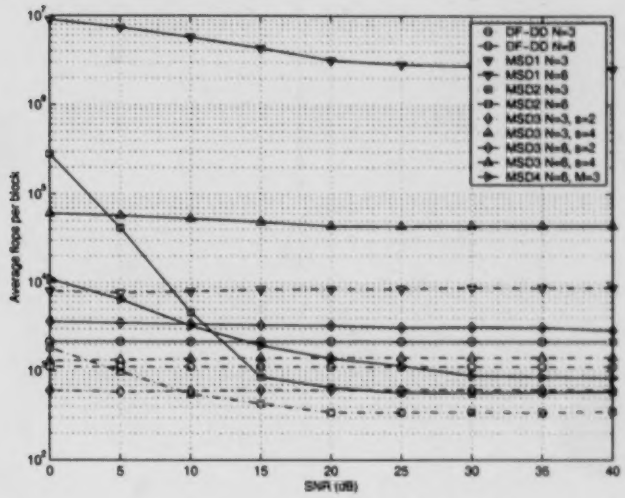

Fig. 4.9. Complexity comparison of $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with $N = 3, 6$ as a function of SNR. The normalized Doppler frequency is $f_d T_s = 0.03$ and $R = 1$.

In Fig. 4.10, we investigate the complexity per block of MSD2 as a function of different N and the normalized Doppler $f_d T_s$ with fixed SNR=30 dB. When $f_d T_s = 0, 0.02$, the complexity per block increases almost linearly as N increases. Different $f_d T_s$ results in different slopes, which is also due to the bound variation by coefficients $\tilde{a}_{i,j}$. When the normalized Doppler frequency is as high as 0.03, the slope is large at first and then becomes flat with increasing N. Fig. 4.10 also suggests higher $f_d T_s$ will cause higher complexity.

4.4.2 Ricean channels

We now show the results over Ricean fading channels. $N_T = 4$, $N_R = 1$ and rate $R = 1$ DUSTM is used as an example.

Fig. 4.11 shows the BER versus SNR for SD-BID, MSD based DF-DD (DF-DD), with $N = 3, 6$, $f_D T = 0.0075$ and Rice factor $K = 5$ dB [70]. Compared with DF-

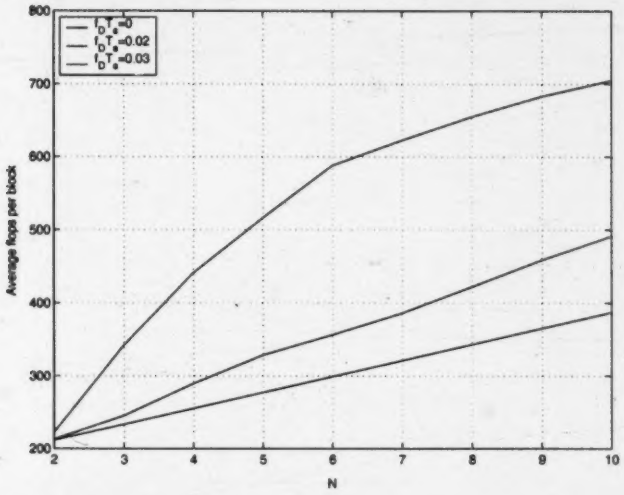

Fig. 4.10. Complexity of MSD2 for $N_T = 4$ transmitter antennas, $N_R = 1$ receiver antenna with SNR=30 dB, $R = 1$ and different normalized Doppler frequencies as a function of N.

DD, the SD-BID has 0.1 dB gain ($N = 3$) and 0.4 dB gain ($N = 6$) at BER= 10^{-4}, respectively. Both DF-DD and SD-BID can reduce the performance gap between CDD and CD. The performance loss of SD-BID over CD is reduced with the increase of N. When $f_D T_B$ increases to 0.03, the gap between SD-BID and CD enlarges from 1 dB to 2 dB (Fig. 4.12). At BER=5×10^{-4}, the DF-DD performs 0.6 dB and 1.2 dB worse than SD-BID. This agrees with the conclusion for SISO systems that the gap between SD-BID and DF-DD enlarges with the increase of N. We also show the performance of RS-DD in Fig. 4.12. When $N = 6$, $M = 3$, RS-DD has about a 0.6-dB gain over SD-BID with $N = 3$, both with 3 dimension's exhaustive search. It outperforms SD-BID by 0.2 dB when $N = 9$, $M = 3$. RS-DD is a good candidate to achieve good performance while still maintain reasonable complexity.

Fig. 4.13 presents the performance of SD-BID ($N = 6$), CDD and CD with different Rice factor K. All the detectors perform better increasing K. The gap between SD-BID and CD reduces with the increase of K and so does the gap between SD-BID and CDD.

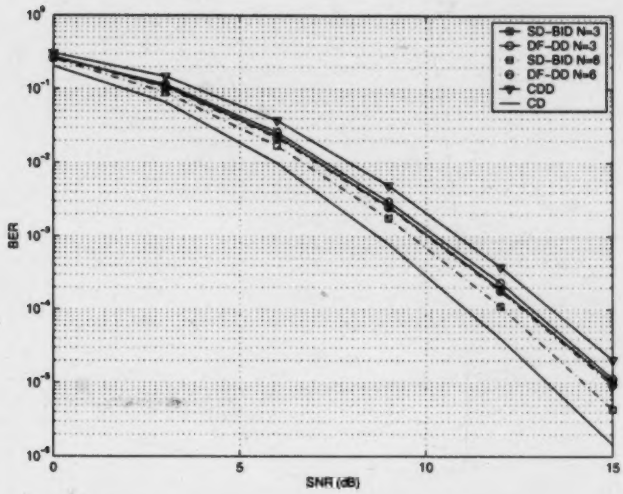

Fig. 4.11. The performance comparison between SD-BID, MSD based DF-DD, CDD and CD with $N = 3, 6$ for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.0075$ and $K = 5$ dB).

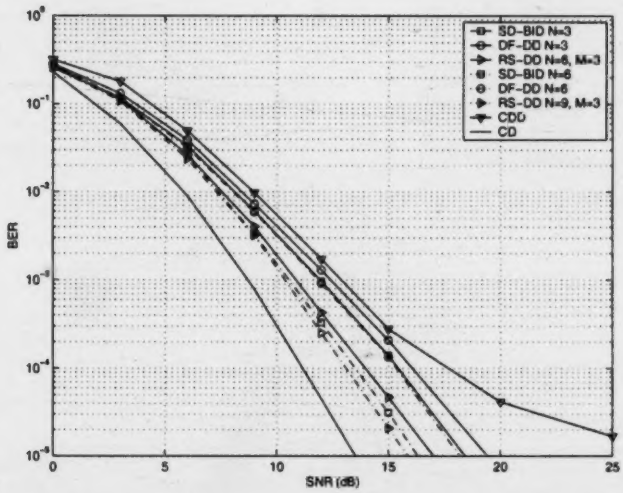

Fig. 4.12. The performance comparison between SD-BID, MSD based DF-DD, CDD and CD with $N = 3, 6$ for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$ and $K = 5$ dB).

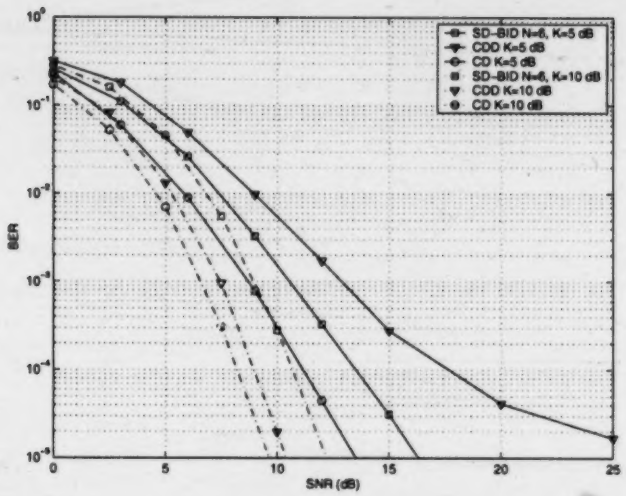

Fig. 4.13. The performance comparison between SD-BID ($N = 6$), CDD and CD for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$) with different Rice factor K.

We compare the complexity of different algorithm in Fig. 4.14. We apply BID to the CD. The complexity of SD-BID reduces as SNR increases. SD-BID is even less complex than DF-DD in the high SNR region. In this region, SD-BID has both complexity and performance gains. Interestingly, the complexity of RS-DD is less than that of SD-BID. In RS-DD, the matrix $\mathbf{U}$ after deleting the corresponding columns to the feedback signals is different from the $\mathbf{U}$ in SD-BID with the same size. The diagonal terms of the matrix in RS-DD is larger than that in pure SD-BID. The structure difference on the matrix offers more complexity savings.

Fig. 4.15 shows the complexity of SD-BID with the increase of K and different $f_D T_B$'s. With the increase of K, the complexity of SD-BID reduces significantly as SD-BID becomes CD when $K \to \infty$. We can also see from Fig. 4.15 that smaller Doppler frequency results in less complexity. Like the argument in RS-DD, the complexity reductions with K and $f_D T_B$ are both because of the change on the structure of matrix $\mathbf{U}$. It is a good example to show the effect of matrix structure in solving CVP.

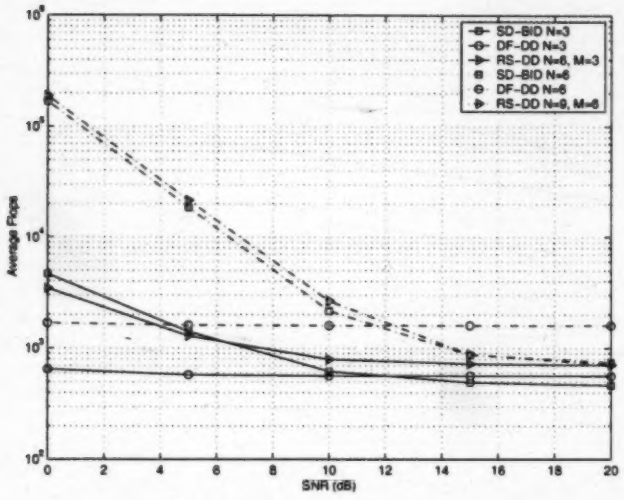

Fig. 4.14. The average number of flops comparison between SD-BID, RS-DD, MSD based DF-DD for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels ($f_D T_B = 0.03$ and $K = 5$ dB).

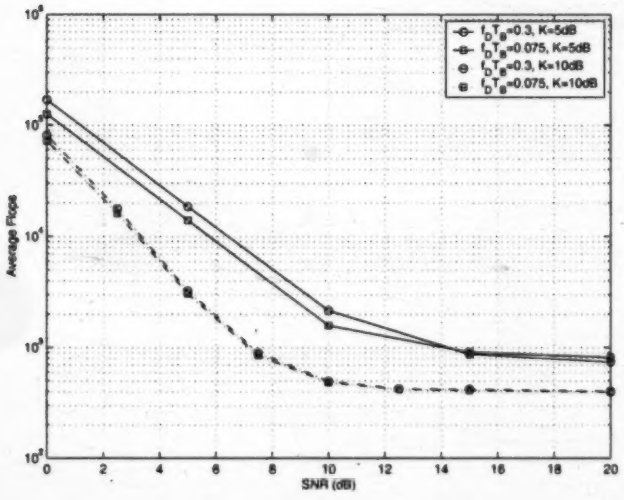

Fig. 4.15. The average number of flops of SD-BID for DUSTM ($N_T = 4$, $N_R = 1$ and $R = 1$) over flat Ricean channels for different $f_D T_B$ and Rice factor K.

4.5 Conclusion

In this chapter, we have considered efficient algorithms for multiple symbol detection of DUSTM over QS fading channels. The multiple symbol detection decision metric for both Rayleigh and Ricean channels are derived. We have derived a novel detection algorithm called BID for single symbol detection of diagonal constellations. This algorithm is exact ML and substantially saves complexity, particulary in high SNR. As well, an interesting and novel feature is the use of the extended Euclidean algorithm for detection. For MSD over Rayleigh channels, we developed four detectors, all of which are derivatives of the BID algorithm. MSD1 and MSD2 are both ML. MSD3 first generates a candidate subset for each l via BID and exhaustively searches over the reduced space. MSD4 generalizes the DF-DD. For the detection of DUSTM over Ricean channels, we have proposed a SD-BID to efficiently solve the MSD. Several efficient implementation issues were also addressed. Simulation results confirm the relationship between Ricean fading, Rayleigh fading and perfect CSI cases.

Chapter 5

Blind and Semi-Blind Data Detection for OFDM Systems

In this chapter, we develop new blind and semi-blind data detectors for Orthogonal Frequency Division Multiplexing (OFDM) systems. Section 5.1 describes the basic baseband OFDM system model. Section 5.2 derives the blind and semi-blind data detectors over frequency selective channels. In Section 5.3, we present efficient detection algorithms for our detectors. We develop a cyclic prefix based power delay profile (PDP) and noise variance estimation algorithm in Section 5.4. Section 5.5 generalizes our data detectors. Section 5.6 shows the simulation results and Section 5.7 concludes this chapter.

5.1 Introduction

5.1.1 Background

OFDM is used for high data rate wireless local area network (WLAN) standards, such as the Hiperlan and IEEE 802.11a, providing data rates of up to 54Mbit/s, and considered for the fourth-generation (4G) mobile wireless systems and beyond [71]. The use of pilot tones for channel estimation [72–74] constitutes a significant overhead or bandwidth loss, motivating the development of blind techniques for OFDM. They use statistical or deterministic properties of the transmit and receive signals, properties such as cyclic prefix (CP) and pilot induced redundancy, cyclostationarity, finite alphabets and virtual carriers have been exploited [75–78].

Joint estimation of channel impulse response (CIR) and data symbols for OFDM has not been investigated extensively. A maximum likelihood (ML) joint blind channel and data estimator [79] exploits the finite alphabet property of modulation symbols and the presence of virtual carriers (VCs). In [80], a blind channel estimator for block fading channels is proposed using the super-trellis and the per-survivor algorithm, which requires relatively high complexity. Recently, [81] also proposed a blind joint channel and data estimator. The branch-and-bound principle is applied to solve a nonlinear integer problem associated with finding the curve that fits a subchannel in the least squares (LS) sense.

Note that many previous blind estimators typically use averaging over large number of OFDM symbols (up to several thousands in some cases). These estimators thereby introduce a considerable latency into the overall system and they also require that the channel remains constant. Thus, estimators that require few OFDM symbols are preferable, as they can operate over non-zero Doppler channels and do not introduce an appreciable delay. The initial channel estimate can also be used for data detection over several OFDM symbols (provided the channel variation is slow

enough).

In this chapter, we develop new blind and semi-blind data detectors for OFDM systems. For frequency selective channels, the semi-blind detector uses both channel correlation and noise variance. The quadratic for the blind detector suffers from rank deficiency, to which we give an efficient solution. Both the detectors are obtained by posing the problem of the joint estimation of channel and data as a mixed discrete and continuous LS optimization problem. By eliminating the channel from it, we obtain a discrete integer LS problem (which has the same form for both the detectors) for the data symbols. Exhaustive search of the solution space yields the ML solution but has exponential complexity in the number of subcarriers and is computationally prohibitive. Avoiding this problem, we solve our data detectors using sphere decoding (SD) [2] and Vertical Bell Laboratories Layered Space-Time (V-BLAST) [1] and provide simple adaptations of the SD algorithm. Our approach allows for substantial computational saving over exhaustive search. Since the semi-blind detector requires both channel correlation and noise variance, we propose a PDP and noise variance estimation algorithm. We also consider how the semi-blind detector performs under mismatch, generalize the basic data detectors to non-unitary constellations and extend them to systems with pilots and virtual carriers. An enhanced data detector is also be derived by noting that for a given LS channel estimate, the true CIR can be modeled as complex Gaussian with mean being the LS channel estimate itself. The LS channel estimate thus gives a prior on the true channel and averaging the likelihood function over the prior distribution gives the enhanced detector that mitigates the effect of channel estimation errors.

5.1.2 OFDM baseband model

In an OFDM system, the source data are grouped and/or mapped into the symbols from a constellation $\mathcal{Q}$, which are modulated by inverse discrete Fourier transform

(IDFT) on N parallel subcarriers. The resulting time domain samples are

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j(2\pi kn/N)}, \quad n = 0, 1, \cdots, N-1 \tag{5.1}$$

where

$$X_k = \begin{cases} d_k & k \in I_d \\ p_k & k \in I_p \end{cases} \tag{5.2}$$

and I_d is the index set of data subcarriers with N_d elements, I_p is the index set of subcarriers reserved for pilot symbols (pilots for brevity) with N_p elements and $N_d + N_p = N$. Note that X_k, $k = 0, 1, \ldots, N-1$ are called OFDM input symbols. We assume that all the d_k's have the same power, i.e., $E\{|d_k|^2\} = E_s$ and all the $E\{|p_k|^2\} = E_p$. The symbols after IDFT is denoted as x_n, $n = 0, \ldots, N-1$. The term "OFDM symbol" denotes the entire IDFT output $\{x_0, x_1, \cdots, x_{N-1}\}$. The input symbol duration is T_s and the OFDM symbol duration is NT_s. These samples are appropriately pulse shaped to construct the time domain signal $x(t)$ for transmission. Typically, pilots X_k, $k \in I_p$, known *a priori* at the receiver, remain fixed from one OFDM symbol to the next. In this pilot arrangement, $N_p \ll N$. Alternatively, entire OFDM symbols of pilots ($N_p = N$) can be transmitted periodically. We focus on the former, as it is more common in applications.

We assume that the composite CIR which includes transmit and receive pulse shaping and the physical channel response between the transmitter and receiver may be modeled as [29, p.802]

$$h(\tau) = \sum_{l=0}^{L-1} h_l \delta(\tau - \tau_l) \tag{5.3}$$

where $h_l \sim \mathcal{CN}(0, \sigma_l^2)$, τ_l is the delay of the lth tap and L is the total number of paths. The $[\sigma_0^2, \cdots, \sigma_{L-1}^2]$ and $[\tau_0, \cdots, \tau_{L-1}]$ constitute the PDP. The received signal after sampling is given by

$$y_n = \sum_{l=0}^{L-1} h_l x_{n-d_l} + w_n \tag{5.4}$$

where $w_n \sim \mathcal{CN}(0, \sigma_n^2)$ is an Additive White Gaussian Noise (AWGN), and $d_l = \lfloor \tau_l / T_s \rceil$ is the delay normalized by T_s. For simplicity, we round d_l to an integer

without considering leakage. However, the detectors in this chapter may also be extended to fractional d_l. We assume perfect synchronization and that the channel remains constant during each OFDM symbol, but it varies between OFDM symbols. If the cyclic prefix is sufficiently long ($N_g > L$), the post-DFT received samples Y_k are given as follows:

$$Y_k = H_k X_k + W_k, \quad 0 \leq k \leq N-1 \tag{5.5}$$

where $H_k = H(j2\pi k/N)$ is the complex channel frequency response at subcarrier k, $H(j\omega)$ is the Fourier transform of the CIR and W_k $k = 0, 1, \cdots, N-1$ is the Fourier transform of w_n and is independent and identically distributed (i.i.d) complex Gaussian random variable (CGRV), each of which also has zero mean and variance σ_n^2. Assuming $\tau_l = lT_s$, we find $\mathbf{H} = \mathbf{F}_L \mathbf{h}$, where $\mathbf{H} = [H_0, H_1, \cdots, H_{N-1}]^T$, $\mathbf{h} = [h_0, h_1, \cdots, h_{L-1}] \in \mathcal{C}^L$ is the CIR and $\mathbf{F}_L$ is a $N \times L$ submatrix of DFT matrix $\mathbf{F}$, which corresponds to each channel path. We can vectorize (5.5) as

$$\mathbf{Y} = \mathbf{X}_D \mathbf{F}_L \mathbf{h} + \mathbf{W} \tag{5.6}$$

where $\mathbf{X}_D = \text{diag}\{X_0, X_1, \cdots, X_{N-1}\}$ is a diagonal matrix. Note that (5.6) is the basis of our blind and semi-blind data detectors.

5.2 Blind and Semiblind Data Detectors

Using ML principles, we derive blind and semi-blind joint CIR estimators and data detectors. The semi-blind detector is derived by assuming the availability of the exact knowledge of channel correlation matrix and noise variance. Our use of the term semi-blind is somewhat unconventional. Typically, semi-blind refers to the use of one or more pilots. We use the term semi-blind to indicate that the detector needs the knowledge of channel correlation and noise variance.

5.2.1 Blind detector

Since W_k's in (5.6) are independent and identically distributed (i.i.d) Gaussian variables, the ML estimator of the channel ($\mathbf{h}$) and transmitted symbols ($\mathbf{X}_D$) is given by

$$(\hat{\mathbf{h}}, \hat{\mathbf{X}}_D) = \underset{\tilde{\mathbf{h}} \in \mathcal{C}^L, \tilde{\mathbf{X}}_D \in \mathcal{Q}^N}{\arg\min} \|\mathbf{Y} - \tilde{\mathbf{X}}_D \mathbf{F}_L \tilde{\mathbf{h}}\|^2. \tag{5.7}$$

The minimization in (5.7) is a complex LS problem for $\hat{\mathbf{h}}$ and an integer LS problem for $\hat{\mathbf{X}}_D$. Given $\hat{\mathbf{X}}_D$ (we assume that $\hat{\mathbf{X}}_D = \mathbf{X}_D$) the channel response $\hat{\mathbf{h}}$ that minimizes (5.7) is given by the LS estimate

$$\hat{\mathbf{h}} = \left[(\hat{\mathbf{X}}_D \mathbf{F}_L)^H (\hat{\mathbf{X}}_D \mathbf{F}_L)\right]^{-1} (\hat{\mathbf{X}}_D \mathbf{F}_L)^H \mathbf{Y}. \tag{5.8}$$

Substituting (5.8) into (5.7), we obtain

$$\hat{\mathbf{X}}_D = \underset{\mathbf{X}_D}{\arg\min} \left\| \mathbf{Y} - \mathbf{X}_D \mathbf{F}_L \left[(\mathbf{X}_D \mathbf{F}_L)^H (\mathbf{X}_D \mathbf{F}_L)\right]^{-1} (\mathbf{X}_D \mathbf{F}_L)^H \mathbf{Y} \right\|^2 \tag{5.9a}$$

$$= \underset{\mathbf{X}_D}{\arg\min} \left\| \left[\mathbf{I}_N - \mathbf{X}_D \mathbf{F}_L \left(\mathbf{F}_L^H \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_D^H\right] \mathbf{Y} \right\|^2 \tag{5.9b}$$

$$= \underset{\mathbf{X}_D}{\arg\min}\, \mathbf{Y}^H \left[\mathbf{I}_N - \mathbf{X}_D \mathbf{F}_L \left(\mathbf{F}_L^H \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_D^H\right] \mathbf{Y} \tag{5.9c}$$

$$= \underset{\mathbf{x}}{\arg\min}\, \mathbf{x}^T \mathbf{Y}_D^H \left[\mathbf{I}_N - \mathbf{F}_L \mathbf{F}_L^H\right] \mathbf{Y}_D \mathbf{x}^* \tag{5.9d}$$

where $\mathbf{Y}_D = \mathrm{diag}\{Y_0, Y_1, \cdots, Y_{N-1}\}$ and $\mathbf{x} \in \mathcal{Q}^N$ is the vector whose elements are the diagonal elements of matrix $\mathbf{X}_D$. Eq.(5.9b) is due to the use of the constant modulus constellation M-PSK; Eq.(5.9c) follows from the fact that the matrix $\mathbf{I}_N - \mathbf{X}_D \mathbf{F}_L \left(\mathbf{F}_L^H \mathbf{F}_L\right)^{-1} \mathbf{F}_L^H \mathbf{X}_D^H$ is an orthogonal projection matrix onto $null(\mathbf{X}_D \mathbf{F}_L)$ and the projection matrix has the property $P_\perp^2 = P_\perp$ and $P_\perp^H = P_\perp$.

The rank of the matrix $\mathbf{B} = \mathbf{Y}_D^H (\mathbf{I}_N - \mathbf{F}_L \mathbf{F}_L^H) \mathbf{Y}_D$ is only $N - L$. Note that $\mathbf{B}$ can be QR factorized as $\mathbf{B} = \mathbf{QR}$, where $\mathbf{Q}$ and $\mathbf{R}$ are unitary and upper-triangular, respectively. Since the last L rows of $\mathbf{R}$ are zero, both the standard V-BLAST and SD algorithms fail here. We next modify (5.9d) so that both SD and V-BLAST can

be applied. Note that in [79], an approximate iterative LS projection algorithm is developed to solve optimization problems similar to (5.9d). However, the convergence of that algorithm is not guaranteed.

Using the constant modulus property (e.g., $|X_k|^2 = 1$ for $X_k \in \mathcal{Q}$), $\mathbf{x}^T\mathbf{Y}_D^H\mathbf{Y}_D\mathbf{x}^* = \sum_{k=0}^{N-1}|Y_k|^2$ is a constant. Therefore the optimization problem (5.9d) is equivalent to

$$\begin{aligned}\hat{\mathbf{X}}_D &= \arg\min_{\mathbf{x}\in\mathcal{Q}^N} \eta\mathbf{x}^T\mathbf{Y}_D^H\mathbf{Y}_D\mathbf{x}^* + \mathbf{x}^T\mathbf{B}\mathbf{x}^* \\ &= \arg\min_{\mathbf{x}\in\mathcal{Q}^N} \mathbf{x}^T\mathbf{Y}_D^H[(\eta+1)\mathbf{I}_N - \mathbf{F}_L\mathbf{F}_L^H]\mathbf{Y}_D\mathbf{x}^*. \end{aligned} \tag{5.10}$$

Since $\mathbf{F}_L\mathbf{F}_L^H$ is positive semi-definite matrix with non-zero eigenvalue 1, $(\eta+1)\mathbf{I}_N - \mathbf{F}_L\mathbf{F}_L^H$ is a positive definite matrix if $\eta > 0$. For simplicity, we let $\eta = \sigma_n^2$.

Detector (5.10) can be solved via an exhaustive search over all M^N possible data sequences, a search whose complexity is exponential in N and which is prohibitive for all but small N. Therefore, we can use both V-BLAST (Section 1.2.1) and SD (Section 1.2.2). Both the algorithms exploit the Cholesky factorization of a positive definite matrix, which can be used for the blind detector (5.10) and the semi-blind detectors developed in the next section.

Remarks:

- The blind detector (5.10) is known as the generalized likelihood ratio test (GLRT) [82]. Similar approach has been used for joint ML channel estimation and signal detection for single input and multiple output systems in [83].
- After $\hat{\mathbf{X}}_D$ is estimated from (5.10), the LS estimate $\hat{\mathbf{h}}$ can be obtained by substituting $\hat{\mathbf{X}}_D$ into (5.8).
- Both $\mathbf{x}$ and $\mathbf{x}e^{j\phi}$, where $\phi \in (0, 2\pi)$, satisfy (5.10), which shows that the blind detector (5.10) exhibits a phase ambiguity. This can be solved by using a pilot tone.

- Existing OFDM standards such as the IEEE802.11a incorporate pilot symbols [84]. These pilots can be used to reduce the search space and solve the phase ambiguity.

5.2.2 Semi-blind detector

This requires the knowledge of the autocorrelation matrix $\mathbf{R}_h$ of the CIR $\mathbf{h}$ and the noise variance σ_n^2. We classify it as a semi-blind detector. From (5.6), $\mathbf{h}$ and $\mathbf{W}$ are zero-mean complex Gaussian random vectors. The received samples Y_k are also zero-mean CGRV's conditioned on $\mathbf{X}_D$. The autocorrelation matrix of the received signal is given by

$$\begin{aligned}\mathbf{R}_Y &= E\{\mathbf{Y}\mathbf{Y}^H\} = \mathbf{X}_D\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_N \\ &= \mathbf{X}_D(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2\mathbf{I}_N)\mathbf{X}_D^H. \end{aligned} \tag{5.11}$$

The determinant of $\mathbf{R}_Y$ can be expressed as

$$\det(\mathbf{R}_Y) = \det(\mathbf{X}_D)\det(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2\mathbf{I}_N)\det(\mathbf{X}_D^H) = \det(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2\mathbf{I}_N). \tag{5.12}$$

Note that the determinant of $\mathbf{R}_Y$ is independent of $\mathbf{X}_D$ if X_k is from a unitary constellation. Ignoring terms that are independent of $\mathbf{X}_D$, the log-likelihood function is given by

$$\Lambda(\mathbf{Y}|\mathbf{X}_D) = -\mathbf{Y}^H\mathbf{R}_Y^{-1}\mathbf{Y}. \tag{5.13}$$

As with (5.9d), maximizing the log likelihood function is equivalent to solving

$$\hat{\mathbf{X}}_D = \arg\min_{\mathbf{x}\in\mathcal{Q}^N} \mathbf{x}^T\mathbf{Y}_D^H(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2\mathbf{I}_N)^{-1}\mathbf{Y}_D\mathbf{x}^*. \tag{5.14}$$

Eq. (5.14) also results in a quadratic form in $\mathbf{x}$.

Remarks:

1. Semi-blind data detection (5.14) also incurs phase ambiguity, as both $\mathbf{x}$ and $\mathbf{x}e^{j\theta}$ satisfy (5.14) if $e^{j\theta}$ belongs to the M-PSK constellation. Pilot symbols are thus needed and the search space hence reduces from $\mathcal{Q}^N$ to $\mathcal{Q}^{N_d}$.

2. Semi-blind data detection (5.14) need not be performed every symbol. If the channel remains constant for M symbols, the channel estimate $\hat{\mathbf{h}}$ obtained in the first symbol using (5.8) can be used to detect the data symbols in the remaining $M-1$ symbols.

3. The semi-blind data detector needs the knowledge of $\mathbf{R}_h$ and σ_n^2, which may not be known exactly. The resulting mismatch problem is studied in Section 5.5.1.

5.2.3 Enhanced data detector with channel estimation error

When the LS channel estimator is used with only a few pilots, the estimated CIR becomes

$$\hat{\mathbf{h}} = \left(\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p\right)^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{Y}_p = \mathbf{h} + \tilde{\mathbf{h}} \tag{5.15}$$

where $\mathbf{P}_D = \text{diag}\{p_1, \ldots, p_{N_p}\}$, $\mathbf{Y}_p = \mathbf{Y}(I_p)$, $\mathbf{F}_p = \mathbf{F}_L(I_p, :)$ is the $N_p \times L$ submatrix of $\mathbf{F}$ corresponding to the pilots and $\tilde{\mathbf{h}} = \left(\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p\right)^{-1} \mathbf{F}_p^H \mathbf{P}_D^H \mathbf{W}$. We can obtain

$$\mathbf{R}_{\tilde{h}} = E\{\tilde{\mathbf{h}}\tilde{\mathbf{h}}^H\} = \sigma_n^2 \left(\mathbf{F}_p^H \mathbf{P}_D^H \mathbf{P}_D \mathbf{F}_p\right)^{-1} = \frac{\sigma_n^2}{E_p} \left(\mathbf{F}_p^H \mathbf{F}_p\right)^{-1} \tag{5.16}$$

where E_p is the power of pilot symbols and we assume that all the pilots have the same energy. If $\hat{\mathbf{h}}$ is used to detect the data in the consecutive OFDM symbols with one-tap equalization, performance loss can be high. This motivates enhanced data detection given channel estimation errors.

Given the *a priori* channel estimate $\hat{\mathbf{h}}$, the true CIR $\mathbf{h}$ has *a priori* Gaussian distribution as

$$p(\mathbf{h}) = \frac{1}{\pi^L \det(\mathbf{R}_{\tilde{h}})} \exp\left\{-(\mathbf{h}-\hat{\mathbf{h}})^H \mathbf{R}_{\tilde{h}}^{-1} (\mathbf{h}-\hat{\mathbf{h}})\right\}. \tag{5.17}$$

The received symbol vector $\mathbf{Y}$ (5.6) is also Gaussian but with mean $\mathbf{X}_D \mathbf{F}_L \mathbf{h}$ and covariance matrix $\sigma_n^2 \mathbf{I}_N$. The likelihood function for the unknown CIR $\mathbf{h}$ and $\mathbf{X}_D$ is

given by

$$f(\mathbf{Y}|\mathbf{h},\mathbf{X}_D) = \exp\left\{-\frac{1}{\sigma_n^2}\|\mathbf{Y}-\mathbf{X}_D\mathbf{F}_L\mathbf{h}\|^2\right\}. \tag{5.18}$$

We average the $f(\mathbf{Y}|\mathbf{h},\mathbf{X}_D)$ in (5.18) with respect to $\mathbf{h}$ with (5.17) resulting in the marginal likelihood function $f(\mathbf{Y}|\mathbf{X}_D)$. We next derive the characteristic function of the following quadratic form

$$Q = \|\mathbf{y}-\mathbf{A}\mathbf{x}\|^2 \tag{5.19}$$

where $\mathbf{A} \in \mathcal{C}^{n\times m}$, $\mathbf{y} \in \mathcal{C}^n$ and $\mathbf{x} \sim \mathcal{CN}(\mu,\mathbf{R})$. Since $\mathbf{z} = \mathbf{y}-\mathbf{A}\mathbf{x} \sim \mathcal{CN}(\mathbf{y}-\mathbf{A}\mu, \mathbf{A}\mathbf{R}\mathbf{A}^H)$, using Eq. (B-3-20) in [85, p. 595], the characteristic function of Q is

$$\begin{aligned}\phi(s) =& E_Q\{e^{-sQ}\} = \int \exp\left\{-s\|\mathbf{z}\|^2\right\} p(\mathbf{z})d\mathbf{z}\\ =& \frac{1}{\det\left(\mathbf{I}+s\mathbf{A}^H\mathbf{A}\mathbf{R}\right)} \exp\Big\{-s\mu^H\mathbf{A}^H\left(s\mathbf{A}\mathbf{R}\mathbf{A}^H+\mathbf{I}\right)^{-1}\mathbf{A}\mu - s\mathbf{y}^H\left(s\mathbf{A}\mathbf{R}\mathbf{A}^H+\mathbf{I}\right)^{-1}\mathbf{y}\\ &+s\mathbf{y}^H\left(s\mathbf{A}\mathbf{R}\mathbf{A}^H+\mathbf{I}\right)^{-1}\mathbf{A}\mu + s\mu^H\mathbf{A}^H\left(s\mathbf{A}^H\mathbf{R}\mathbf{A}+\mathbf{I}\right)^{-1}\mathbf{y}\Big\}.\end{aligned} \tag{5.20}$$

Let $Q = \|\mathbf{Y}-\mathbf{X}_D\mathbf{F}_L\mathbf{h}\|^2$. Substituting $\mathbf{A} = \mathbf{X}_D\mathbf{F}_L$, $\mu = \hat{\mathbf{h}}$ and $\mathbf{R} = \mathbf{R}_{\tilde{h}}$ into (5.20), we obtain

$$\begin{aligned}f(\mathbf{Y}|\mathbf{X}_D) = &\frac{1}{\pi^N \det(\sigma_n^2\mathbf{I}_N + \mathbf{F}_L^H\mathbf{X}_D^H\mathbf{X}_D\mathbf{F}_L\mathbf{R}_{\tilde{h}})} \exp\Big\{-\hat{\mathbf{h}}^H\left(\mathbf{R}_{\tilde{h}}+\sigma_n(\mathbf{F}_L^H\mathbf{X}_D^H\mathbf{X}_D\mathbf{F}_L)^{-1}\right)^{-1}\hat{\mathbf{h}}\\ &-\mathbf{Y}^H(\mathbf{X}_D\mathbf{F}_L\mathbf{R}_{\tilde{h}}\mathbf{F}_L^H\mathbf{X}_D^H+\sigma_n^2\mathbf{I}_N)^{-1}\mathbf{Y} + 2\mathrm{Re}\left[\mathbf{Y}^H\left(\mathbf{X}_D\mathbf{F}_L\mathbf{R}_{\tilde{h}}\mathbf{F}_L^H\mathbf{X}_D^H+\sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{X}_D\mathbf{F}_L\hat{\mathbf{h}}\right]\Big\}.\end{aligned} \tag{5.21}$$

If X_k's are unitary, $\det(\sigma_n^2\mathbf{I}_N + \mathbf{F}_L^H\mathbf{X}_D^H\mathbf{X}_D\mathbf{F}_L\mathbf{R}_{\tilde{h}})$ is independent of $\mathbf{X}_D$.

Ignoring the terms independent of $\mathbf{X}_D$, maximizing (5.21) is equivalent to minimizing

$$\begin{aligned}g_x(\mathbf{X}) =& \mathbf{Y}^H(\mathbf{X}_D\mathbf{F}_L\mathbf{R}_{\tilde{h}}\mathbf{F}_L^H\mathbf{X}_D^H+\sigma_n^2\mathbf{I}_N)^{-1}\mathbf{Y} - 2\mathrm{Re}\left[\mathbf{Y}^H\left(\mathbf{X}_D\mathbf{F}_L\mathbf{R}_{\tilde{h}}\mathbf{F}_L^H\mathbf{X}_D^H+\sigma_n^2\mathbf{I}_L\right)^{-1}\mathbf{X}_D\mathbf{F}_L\hat{\mathbf{h}}\right]\\ =& \left\|(\mathbf{F}_L\mathbf{R}_{\tilde{h}}\mathbf{F}_L^H+\sigma_n^2\mathbf{\Lambda}^{-1})^{-\frac{1}{2}}\left(\mathbf{F}_L\hat{\mathbf{h}}-\mathbf{Y}_D\mathbf{X}^*\right)\right\|^2\end{aligned} \tag{5.22}$$

where Λ is a diagonal matrix with

$$[\mathbf{\Lambda}]_{k,k} = \begin{cases} E_s & k \in I_d \\ E_p & k \in I_p \end{cases}. \tag{5.23}$$

Substituting (5.16) into (5.22), we have

$$g_x(\mathbf{X}) = \left\| (\frac{\sigma_n^2}{E_p}\mathbf{F}_L \left(\mathbf{F}_p^H\mathbf{F}_p\right)^{-1}\mathbf{F}_L^H + \sigma_n^2\mathbf{\Lambda}^{-1})^{-\frac{1}{2}} \left(\mathbf{F}_L\hat{\mathbf{h}} - \mathbf{Y}_D\mathbf{X}^*\right)\right\|^2$$
$$= \left\| \frac{E_p}{\sigma_n^2}(\mathbf{F}_L \left(\mathbf{F}_p^H\mathbf{F}_p\right)^{-1}\mathbf{F}_L^H + E_p\mathbf{\Lambda}^{-1})^{-\frac{1}{2}} \left(\mathbf{F}_L\hat{\mathbf{h}} - \mathbf{Y}_D\mathbf{X}^*\right)\right\|^2. \qquad (5.24)$$

Since the scalar E_p/σ_n^2 does not affect the minimum of (5.24), the enhanced data detector is now given by

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}\in\mathcal{Q}^N} \left\|(\mathbf{F}_L \left(\mathbf{F}_p^H\mathbf{F}_p\right)^{-1}\mathbf{F}_L^H + E_p\mathbf{\Lambda}^{-1})^{-\frac{1}{2}} \left(\mathbf{F}_L\hat{\mathbf{h}} - \mathbf{Y}_D\mathbf{X}^*\right)\right\|^2. \qquad (5.25)$$

Remarks:

1. Unlike (5.15), a remarkable advantage of the enhanced detector (5.24) is that it requires neither $\mathbf{R}_h$ nor σ_n^2. The only required parameter is the pilot energy E_p, which is known at the receiver. Therefore, no parameter mismatch problem exists in (5.24).

2. The performance gain of (5.24) is obtained by using the statistics of the channel estimation error and sacrificing the simplicity of one-tap equalization. When more computational power is available, (5.24) saves the transmission bandwidth and power, providing a tradeoff between complexity and bandwidth efficiency.

5.3 Data Detection Algorithms

The blind detector (5.10) and the semi-blind detector (5.14) can both be written in a general form as

$$\hat{\mathbf{X}}_D = \arg\min_{\mathbf{x}\in\mathcal{Q}^N} \mathbf{x}^T\mathbf{G}\mathbf{x}^* = \|\mathbf{M}\mathbf{x}^*\|^2 \qquad (5.26)$$

where $\mathbf{G}$ is a positive definite matrix, which can be Cholesky factored as $\mathbf{G} = \mathbf{M}^H\mathbf{M}$. Eq. (5.26) is the same as the data detection problem in (1.2) given $\mathbf{r} = 0$ and $\mathbf{M} = \mathbf{H}$.

Similarly, the enhanced data detector (5.25) can also be written in the form (1.2). Therefore both detectors can be solved via V-BLAST and SD. Details can be found in Sections 1.2.1, 1.2.2.

In Section 1.2, we show that a system employing complex constellations such as square QAM can be transformed in to a PAM modulated system. However not all constellations can be decoupled into real systems (e.g., 8-PSK). In [30], a modified sphere decoder is proposed to handle PSK constellations. But it involves the computationally inefficient $\cos^{-1}$ operation, slowing down the SD. We address here the decouple algorithm can still be used to handle M-PSK and any other constellations. Note that each element of $\mathbf{x}$ is constrained by the constellation. For a given $\Re(x_k)$, the candidates for $\Im(x_k)$ are hence constrained. Let $\mathcal{Q}_R = \{\Re(x)|x \in \mathcal{Q}\}$. Let $\mathcal{Q}_I(x) = \{\Im(x)|x \in \mathcal{Q}, x \in \mathcal{Q}_R\}$. Therefore x_k is selected from

$$x_k \in \begin{cases} [LB_k, UB_k] \cap \mathcal{Q}_R & k = 2i, i = 0, 1, \ldots, N-1 \\ [LB_k, UB_k] \cap \mathcal{Q}_I(x_{k-1}) & k = 2i+1, i = 0, 1, \ldots, N-1 \end{cases} . \tag{5.27}$$

where LB_k and UB_k are defined in Algorithm 1. Since $Q_I(x)$ can be pre-computed for each x from $\mathcal{Q}_R$ and be stored in memory, additional computational complexity is avoided. This simple idea can be used to handle any constellations. The decoupling for M-QAM can be viewed as a special case of our generalization because $Q_I(x)$'s are the same for any x from $\mathcal{Q}_R$ and $\mathcal{Q}_R = \mathcal{Q}_I$.

An important problem is how to choose the initial radius. For OFDM symbols with M-PSK, we relax (5.10) and (5.14) as $\min \mathbf{x}^H\mathbf{G}\mathbf{x}$, $\mathbf{x}^H\mathbf{x} = N$ where the vector $\mathbf{x} \in \mathcal{C}^N$. The Lagrangian $\mathcal{L}(\mathbf{x}, \lambda)$ for this minimization problem is $\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^H\mathbf{G}\mathbf{x} + \lambda(\mathbf{x}^H\mathbf{x} - N)$. The optimal λ here is the maximum eigenvalue of matrix $\mathbf{G}$ and $\tilde{\mathbf{x}}$ is the eigenvector corresponding to λ. We quantize $\tilde{\mathbf{x}}$ into a point in $\mathcal{Q}^N$ as $\hat{\mathbf{x}}$. By substituting $\hat{\mathbf{x}}$ into (5.26), the initial radius is given by $d^2 = \hat{\mathbf{x}}^H\mathbf{G}\hat{\mathbf{x}}$.

5.4 Power Delay Profile and Noise Variance Estimation

The semi-blind data detector (5.14) needs the knowledge of channel autocorrelation matrix and noise variance. The autocorrelation matrix is determined by the PDP. Other than our semi-blind detector, in OFDM systems, noise variance and PDP are needed for many algorithms such as MMSE channel estimation and ML frequency offset estimation. Noise variance estimation is also required in many communication applications such as adaptive modulation, turbo coding and others. In [86], a noise-variance estimator is proposed that directly uses the receiver statistics. A subspace approach is presented in [87] that uses the sample covariance matrix of the received signal. However, both algorithms are data aided (DA) estimators, which constitute a bandwidth loss. The estimation of the number of multipath gains and associated time delays have been proposed in [88], where pilot symbols are also needed, and channel multipath power and noise variance are required. In [89], a noise variance and SNR estimator that uses training symbols is developed for multiple antenna OFDM systems. Except for these contributions, no other non-data aided (NDA) noise variance and PDP estimators for OFDM systems have been published to date.

We next develop cyclic prefix based NDA noise-variance and PDP estimators for OFDM systems over multipath fading channels. The key is to use the fact that the cyclic prefix contains the repeated samples which introduces a special correlation structure on the received samples. The noise variance, the number of multipath taps, and PDP are jointly estimated without pilots. The ML function for the estimated parameters is derived, resulting in an ML estimator.

We consider multiple OFDM symbols in this section. Therefore the OFDM symbol during the mth block interval is denoted as $x_n(m)$ and the received signal is denoted as $y_n(m)$. We assume perfect synchronization, and that the PDP is invariant within M OFDM symbols. If there exists a synchronization error, a decision directed algorithm

may be applied using our proposed parameter estimators and the joint ML time and frequency offset estimator in [90].

At the border between two OFDM blocks ($-N_g \leq n < 0$), the received signal samples can be written as

$$y_n(m) = \sum_{l=0}^{L-1} h_l x_{n-d_l}(m) U(n-d_l) + \sum_{l=0}^{L-1} h_l x_{N+n-d_l}(m-1) U(d_l-n) + w_n(m) \tag{5.28}$$

where $U(\cdot)$ is the step function. The correlation between each received signal sample over the CP interval and its corresponding sample at the end of the OFDM block can thus be given by

$$E\{y_{-k}(m)y^*_{N-k}(m)\} = \begin{cases} \sigma_y^2 + \sigma^2 & 0 < k \leq N_g - d_{L-1} \\ \sum_{l=0}^{L-1} \sigma_l^2 U(N_g - k - d_l) & N_g - d_{L-1} < k \leq N_g - d_0 \\ 0 & N_g - d_0 < k \leq N_g \end{cases} \tag{5.29}$$

where $\sigma_y^2 = \sum_{l=0}^{L-1} \sigma_l^2$, and $k = 1, \ldots, N_g$. The expectation in (5.29) is taken with respect to both h_l and $x_n(m)$.

When L is large, $y_n(m)$ can be modeled **approximately** as complex Gaussian using the central limit theorem, and the probability density function (pdf) is given by

$$f(y_n(m)) = \frac{\exp\left(-\frac{|y_n(m)|^2}{\sigma_y^2+\sigma^2}\right)}{\pi(\sigma_y^2+\sigma^2)}. \tag{5.30}$$

Samples $y_{-k}(m)$ and $y_{N-k}(m)$ are jointly Gaussian with pdf

$$f(y_{-k}(m), y_{N-k}(m)) = \frac{\exp\left(-\frac{|y_{-k}(m)|^2+|y_{N-k}(m)|^2-2\rho_k \Re\{y_{-k}(m)y^*_{N-k}(m)\}}{(\sigma_y^2+\sigma^2)(1-\rho_k^2)}\right)}{\pi^2(\sigma_y^2+\sigma^2)(1-\rho_k^2)} \tag{5.31}$$

where

$$\begin{aligned} \rho_k &= \left| \frac{E\{y_{-k}(m)y^*_{N-k}(m)\}}{\sqrt{E\{|y_{-k}(m)|^2\}E\{|y_{N-k}(m)|^2\}}} \right| \\ &= \frac{\sum_{l=0}^{L-1} \sigma_l^2 U(N_g - k - d_l)}{\sum_{l=0}^{L-1} \sigma_l^2 + \sigma^2}. \end{aligned} \tag{5.32}$$

Therefore, the proposed estimator is only approximate ML.

We use M OFDM blocks to estimate those parameters and assume that they remain unchanged during the M blocks. Define $\mathbf{p} = [\sigma_0^2, \ldots, \sigma_{L-1}^2]$, $\mathbf{d} = [d_0, \ldots, d_{L-1}]$ and $\mathbf{y} = [y_{-N_g}(1), y_{-N_g+1}(1), \ldots, y_{N-1}(M)]$. Using (5.30) and (5.31) and assuming the M OFDM blocks are independent, the log-likelihood function of $\mathbf{y}$ conditioned on $\sigma^2, \mathbf{p}, \mathbf{d}$ can be written as

$$\begin{aligned}
&\Lambda(\mathbf{y}|\sigma^2, \mathbf{p}, \mathbf{d}) \\
&= \sum_{m=1}^{M} \log \left(\prod_{k=1}^{N_g} f(y_{-k}(m), y_{N-k}(m)) \prod_{k=0}^{N'} f(y_k(m)) \right) \\
&= -M \left(\sum_{k=1}^{N_g} \frac{a_k - \rho_k b_k}{c(1-\rho_k^2)} + \log(c(1-\rho_k^2)) + \sum_{k=0}^{N'} \frac{g_k}{c} + \log(c) \right) \qquad (5.33)
\end{aligned}$$

where $N' = N - N_g - 1$,

$$\begin{aligned}
a_k &= \frac{\sum_{m=1}^{M} |y_{-k}(m)|^2 + |y_{N-k}(m)|^2}{M} \\
b_k &= \frac{\sum_{m=1}^{M} \Re\{y_{-k}(m) y_{N-k}^*(m)\}}{M} \\
g_k &= \frac{\sum_{m=1}^{M} |y_k(m)|^2}{M}, \; c = \sigma_y^2 + \sigma^2. \qquad (5.34)
\end{aligned}$$

Since (5.33) involves many variables, to simplify the joint parameters' estimation, we take a suboptimal way. We first estimate c by maximizing only the last sum in (5.33):

$$\hat{c} = \frac{\sum_{k=0}^{N'} g_k}{N - N_g} = \frac{\sum_{k=0}^{N'} \sum_{m=1}^{M} |y_k(m)|^2}{(N - N_g)M}. \qquad (5.35)$$

From (5.35), we find $\hat{c}$ is the time average estimation of $\sigma_y^2 + \sigma^2$, and hence an estimate of c is given by $\hat{c}$. Substituting $\hat{c}$ back into the first summation of (5.33) and maximizing ρ_k individually, we get the estimate for ρ_k as the real root of the equation

$$2\hat{c}\rho^3 - b\rho^2 - 2(\hat{c} - a_k)\rho - b = 0. \qquad (5.36)$$

We then compute the value s_k as

$$s_k = \begin{cases} \rho_{N_g} \hat{c} & k = 1 \\ (\rho_{N_g - k + 1} - \rho_{N_g - k + 2})\hat{c} & k = 2, \ldots, N_g \end{cases} . \qquad (5.37)$$

A threshold value is set as $\alpha\hat{c}$, where α is a constant less than 1. If $s_k > \alpha\hat{c}$, it is identified as a path; s_k is the estimate of path power, and k is the estimate of delay time. The number of paths is estimated as the number of s_k that $s_k > \alpha\hat{c}$. We denote the maximum delay time as $d_{\max}$. The noise variance can thus be estimated as

$$\hat{\sigma}^2 = \hat{c}\left(1 - \frac{\sum_{k=1}^{N_g - d_{\max}+1} \hat{\rho}_k}{N_g - d_{\max} + 1}\right). \tag{5.38}$$

If we look directly at the structure of OFDM block, in the absence of noise, $y_{-k}(m) = y_{N-k}(m)$ for $k = 1, \ldots, N_g - d_{\max} + 1$. The noise variance can be obtained alternatively as

$$\hat{\sigma}^2 = \frac{\sum_{m=1}^{M} \sum_{k=1}^{N_g - d_{\max}+1} |y_{N-k}(m) - y_{-k}(m)|^2}{2M(N_g - d_{\max} + 1)}. \tag{5.39}$$

Using the results of (5.35) and (5.38) or (5.39), SNR can be estimated by

$$\text{SNR} = \frac{\hat{c}}{\hat{\sigma}^2}. \tag{5.40}$$

Note that (5.39) can only be used to estimate σ^2.

Note that the SNR considered in this section is average SNR, which is averaged over the channel, data and noise realizations. Our proposed algorithm cannot estimate the instantaneous SNR, where a fixed channel is considered.

We now investigate the performance of our proposed estimators. We assume an OFDM system using QPSK with $N = 64$ subcarriers, and CP length $N_g = 16$. A $L = 6$ channel model is used. The power profile is given by $\mathbf{p} = [0.189, 0.379, 0.239, 0.095, 0.061, 0.037]$, and the delay profile after sampling is $\mathbf{d} = [0, 1, 2, 4, 6, 8]$. Each path is an independent, zero-mean complex Gaussian random process.

Fig. 5.1 shows the probability of correct detection of the number of paths using our proposed algorithm with different M. The threshold parameter is set to $\alpha = 0.01$. In low SNR, the paths with smaller power are dominated by the noise, and there may

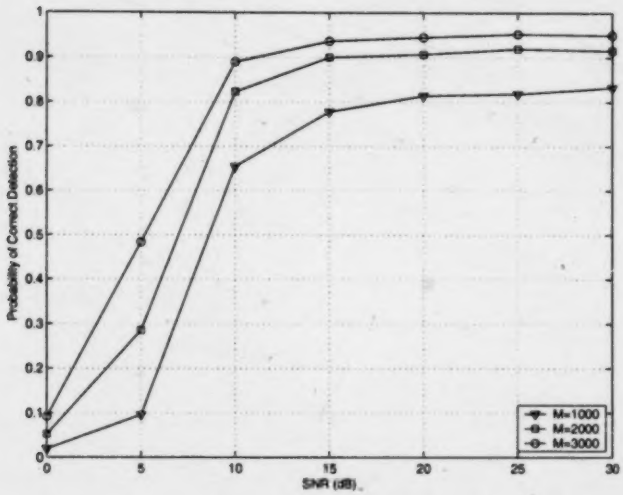

Fig. 5.1. The probability of correct detection of the number of paths.

be many paths larger than the threshold. Therefore, the number of paths may be overestimated. The probability of correct detection increases in high SNR. With increasing M, the probability of detection error decreases.

Fig. 5.2 presents the normalized mean square error (NMSE) of the channel power estimation for the 3rd path (arbitrarily chosen), where the NMSE is defined as NMSE $= E\{(\hat{\sigma}_3^2 - \sigma_3^2)^2\}/\sigma_3^4$. The channel power is overwhelmed by the noise in low SNR. In high SNR, the NMSE becomes constant since the number of paths cannot be 100% correctly detected. The NMSE is improved by increasing M.

Fig. 5.3 shows the NMSE of the noise variance estimation using different estimators, where the NMSE is defined as NMSE $= E\{(\hat{\sigma}^2 - \sigma^2)^2\}/\sigma^4$. The estimator using (5.38) is denoted as ML, and that using (5.39) is denoted as direct estimator or (DI). At low SNR, the DI method performs better than the approximate ML method since the probability of $d_{\max}$ detection is higher for the DI method. In high SNR, both DI and ML perform identically. With the increase of M, the performance of both estimators improve.

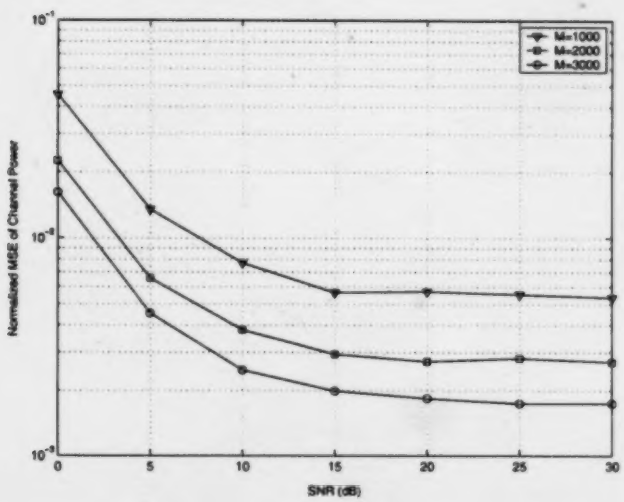

Fig. 5.2. The NMSE of the channel power estimation for the 3rd path.

In this section, we have presented noise-variance and power-delay-profile estimators using the CP in each OFDM block. The correlation structure due to the use of the CP has been exploited to derive our estimators, and hence pilot symbols are not needed. A direct heuristic noise variance estimator has also been proposed. Simulation results show that our proposed estimators provide an effective way to estimate the channel parameters. The results in this section may be used to improve the performance and reduce the complexity of channel estimators for OFDM systems.

5.5 Generalizations

5.5.1 Semi-blind detector under mismatch

Even though we can estimate the channel covariance matrix $\mathbf{R}_h$ and the noise variance σ_n^2 using the algorithm in Section 5.4. There still exists residual mismatch. Hence, we investigate the design of robust semi-blind detectors for $\tilde{\mathbf{R}}_h$ and $\tilde{\sigma}_n^2$, while the true values are $\mathbf{R}_h$ and σ_n^2. In [72, 73], the estimators against the mismatch are designed

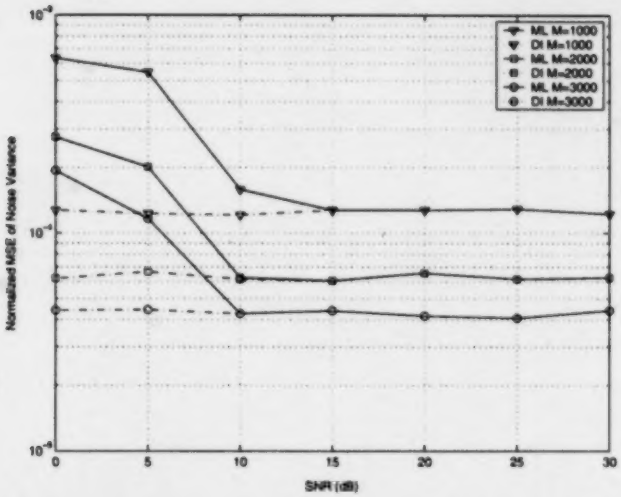

Fig. 5.3. The NMSE of the noise variance estimation using different estimators.

for the worst case, which is taken to be the uniform power delay profile (UPDP). We follow their approach and consider a suboptimal criterion as follows:

$$\begin{aligned}&\mathrm{E}\left\{\mathbf{Y}^H\hat{\mathbf{X}}_D(\mathbf{F}_L\tilde{\mathbf{R}}_h\mathbf{F}_L^H+\tilde{\sigma}_n^2\mathbf{I})^{-1}\hat{\mathbf{X}}_D^H\mathbf{Y}\right\}\\ =&\mathrm{E}\left\{\mathbf{h}^H\mathbf{F}_L^H\mathbf{X}_D^H\hat{\mathbf{X}}_D(\mathbf{F}_L\tilde{\mathbf{R}}_h\mathbf{F}_L^H+\tilde{\sigma}_n^2\mathbf{I})^{-1}\hat{\mathbf{X}}_D^H\mathbf{X}_D\mathbf{F}_L\mathbf{h}+\mathbf{W}^H\hat{\mathbf{X}}_D(\mathbf{F}_L\tilde{\mathbf{R}}_h\mathbf{F}_L^H+\tilde{\sigma}_n^2\mathbf{I})^{-1}\hat{\mathbf{X}}_D^H\mathbf{W}\right\}\end{aligned} \tag{5.41}$$

The second equality comes from (5.6). If the first term of (5.41) is independent of $\tilde{\mathbf{R}}_h$, the BER may be less dependent of the mismatch. Ignoring the noise variance σ_n^2 and letting $\Delta\mathbf{X}=\hat{\mathbf{X}}_D^H\mathbf{X}_D$, the first term of (5.41) can be written as

$$\begin{aligned}&\mathrm{E}\left\{\mathbf{h}^H\mathbf{F}_L^H\mathbf{X}_D^H\hat{\mathbf{X}}_D(\mathbf{F}_L\tilde{\mathbf{R}}_h\mathbf{F}_L^H+\tilde{\sigma}_n^2\mathbf{I})^{-1}\hat{\mathbf{X}}_D^H\mathbf{X}_D\mathbf{F}_L\mathbf{h}\right\}\\ \simeq&\mathrm{E}\left\{\mathbf{h}^H\mathbf{F}_L^H\Delta\mathbf{X}^H(\mathbf{F}_L\tilde{\mathbf{R}}_h\mathbf{F}_L^H)^{-1}\Delta\mathbf{X}\mathbf{F}_L\mathbf{h}\right\}\\ =&\mathrm{E}\left\{\mathbf{h}^H\mathbf{F}_L^H\Delta\mathbf{X}^H\mathbf{F}_L(\tilde{\mathbf{R}}_h)^{-1}\mathbf{F}_L^H\Delta\mathbf{X}\mathbf{F}_L\mathbf{h}\right\}\end{aligned} \tag{5.42}$$

Since $\mathbf{F}_L^H\Delta\mathbf{X}\mathbf{F}_L$ is a circulant matrix, we assume the first row is denoted by $a_0,a_1,\cdots,a(L-1)$. Eq.(5.42) becomes

$$\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{|a((i+j)\mathrm{mod}L)|^2}{\tilde{\mathbf{R}}_h(j)}\mathbf{R}_h(i) \tag{5.43}$$

If $\tilde{\mathbf{R}}_h(j)$ is UPDP or $\tilde{\mathbf{R}}_h(j) = 1/L$, (5.43) is equal to $L\sum_{i=0}^{L-1}|a(i)|^2$ which is independent of $\tilde{\mathbf{R}}_h$. Hence the detector is robust to mismatch with this choice of $\tilde{\mathbf{R}}_h$.

5.5.2 Generalization to non-unitary constellations

The blind and semi-blind detectors can be extended to non-unitary constellations. Instead of solving the optimal detector (5.7), the suboptimal blind detector solves

$$(\hat{\mathbf{h}}, \hat{\mathbf{X}}_D) = \underset{\tilde{\mathbf{h}}\in\mathcal{C}^L, \tilde{\mathbf{X}}_D\in\mathcal{Q}^N}{\arg\min} \|\tilde{\mathbf{X}}_D^{-1}\mathbf{Y} - \mathbf{F}_L\tilde{\mathbf{h}}\|^2 \tag{5.44}$$

where $\mathcal{Q}$ denotes the non-unitary constellation. The LS estimate of $\mathbf{h}$ is given by

$$\hat{\mathbf{h}} = \mathbf{F}_L^H\hat{\mathbf{X}}_D^{-1}\mathbf{Y} \tag{5.45}$$

Substituting (5.45) into (5.44), we obtain

$$\begin{aligned}\hat{\mathbf{X}}_D &= \underset{\mathbf{X}_D}{\arg\min}\, \mathbf{Y}^H\mathbf{X}_D^{-H}\left(\mathbf{I}_N - \mathbf{F}_L\mathbf{F}_L^H\right)\mathbf{X}_D^{-1}\mathbf{Y} \\ &= \underset{\mathbf{x}\in(\mathcal{Q}^{-1})^N}{\arg\min}\, \mathbf{x}^H\mathbf{Y}_D^H\left(\mathbf{I}_N - \mathbf{F}_L\mathbf{F}_L^H\right)\mathbf{Y}_D\mathbf{x}\end{aligned} \tag{5.46}$$

where $\mathcal{Q}^{-1}$ denotes the constellation whose element is the inverse of the corresponding element in $\mathcal{Q}$. Eq. (5.46) can also be solved using V-BLAST and SD but it suffers from performance loss due to the suboptimal detector (5.44).

For the semi-blind detector, we note that (5.11) is

$$\mathbf{R}_Y = \mathbf{X}_D\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_N. \tag{5.47}$$

Maximizing the log likelihood function is equivalent to solving

$$\begin{aligned}\hat{\mathbf{X}}_D &= \underset{\mathbf{X}_D}{\arg\min}\, \mathbf{Y}^H\left(\mathbf{X}_D\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H\mathbf{X}_D^H + \sigma_n^2\mathbf{I}_N\right)^{-1}\mathbf{Y} \\ &= \underset{\mathbf{X}_D}{\arg\min}\, \mathbf{Y}^H\mathbf{X}_D^{-H}\left(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2(\mathbf{X}_D\mathbf{X}_D^H)^{-1}\right)^{-1}\mathbf{X}_D^{-1}\mathbf{Y}.\end{aligned} \tag{5.48}$$

As in [72], we derive a suboptimal detector by replacing the term $(\mathbf{X}_D\mathbf{X}_D^H)^{-1}$ in (5.48) with its expectation $E\{(\mathbf{X}_D\mathbf{X}_D^H)^{-1}\}$. Assuming the same constellation on all

the subcarriers, we have $E\{(\mathbf{X}_D\mathbf{X}_D^H)^{-1}\} = \rho\mathbf{I}_N$, where $\rho = E\{1/|x_k|^2\}$. Therefore the suboptimal semi-blind detector is given by

$$\hat{\mathbf{X}}_D = \arg\min_{\mathbf{x}\in(\mathcal{Q}^{-1})^N} \mathbf{x}^H\mathbf{Y}_D^H(\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \rho\sigma_n^2\mathbf{I}_N)^{-1}\mathbf{Y}_D\mathbf{x}. \tag{5.49}$$

For optimal detection of (5.48), we note that $\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H$ is positive semi-definite. It can be readily verified that $\mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2(\mathbf{X}_D\mathbf{X}_D^H)^{-1}$ is positive definite. We use the following definition from [69, p. 469].

Definition: Let $\mathbf{A}$, $\mathbf{B}$ be Hermitian matrices. We write $\mathbf{A} \succeq \mathbf{B}$ if the matrix $\mathbf{A} - \mathbf{B}$ is positive semi-definite. Similarly, $\mathbf{A} \succ \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive definite.

Let $\mathbf{A} = \mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \sigma_n^2(\mathbf{X}_D\mathbf{X}_D^H)^{-1}$ $\mathbf{B} = \mathbf{F}_L\mathbf{R}_h\mathbf{F}_L^H + \rho_m\sigma_n^2\mathbf{I}_N$, where $\rho_m = \max\{1/|x_k|^2\}$. It can be readily verified that $\mathbf{B} \succeq \mathbf{A}$. Using Corollary [69, p. 471], we can obtain $\mathbf{A}^{-1} \succeq \mathbf{B}^{-1}$. Therefore, for any $\mathbf{x} \in (\mathcal{Q}^{-1})^N$

$$\mathbf{x}^H\mathbf{Y}_D^H\mathbf{B}^{-1}\mathbf{Y}_D\mathbf{x} \leq \mathbf{x}^H\mathbf{Y}_D^H\mathbf{A}^{-1}\mathbf{Y}_D\mathbf{x} \tag{5.50}$$

When applying the SD to (5.48), it find the minimum point among all the points satisfying

$$\mathbf{x}^H\mathbf{Y}_D^H\mathbf{A}^{-1}\mathbf{Y}_D\mathbf{x} \leq r^2. \tag{5.51}$$

Using (5.50), we find the point minimizing (5.48) among all the points satisfying

$$\mathbf{x}^H\mathbf{Y}_D^H\mathbf{B}^{-1}\mathbf{Y}_D\mathbf{x} \leq r^2. \tag{5.52}$$

Hence, the SD should be modified when the search goes to the bottom of the search tree, the SD updates the radius r according to $\mathbf{x}^H\mathbf{Y}_D^H\mathbf{A}^{-1}\mathbf{Y}_D\mathbf{x}$ instead of $\mathbf{x}^H\mathbf{Y}_D^H\mathbf{B}^{-1}\mathbf{Y}_D\mathbf{x}$. This gives the optimal solution.

5.5.3 Time-varying channel tracking via decision feedback

Our proposed detectors may also be used for channel tracking if the channel remains constant for K OFDM symbols. For $p = 1$ (the first symbol), the initial channel estimate $\hat{\mathbf{h}}(1)$ is obtained by blind or semi-blind detectors. For the remaining OFDM symbols ($p = 2, \cdots, K$), channel estimation may not be necessary. Instead, $\hat{\mathbf{h}}(1)$ is used to detect $p = 2, \ldots, K$ OFDM symbols. Decision-feedback type iterations can also be used to track a slowly-varying channel. For the pth symbol, the OFDM symbol is detected using the channel estimate in the $p-1$th symbol and is denoted as $\hat{\mathbf{X}}_D(p)$. The channel in the pth symbol $\mathbf{h}(p)$ is then updated as

$$\hat{\mathbf{h}}(p) = \left[(\hat{\mathbf{X}}_D(p)\mathbf{F}_L)^H(\hat{\mathbf{X}}_D(p)\mathbf{F}_L)\right]^{-1}(\hat{\mathbf{X}}_D(p)\mathbf{F}_L)^H\mathbf{Y}(p) \tag{5.53}$$

where $\mathbf{Y}_p$ denotes the received symbols in the pth symbol.

5.6 Simulation Results

Simulation results are given for the proposed detectors. We consider a frequency-selective slow Rayleigh fading channel with L Gaussian complex taps h_l with $\sigma_l^2 = E[|h_l|^2] = \sigma_0^2 e^{-l/5}$ for $l = 1, \cdots, L$; and the six-tap COST 207 TU channel model [91], which has the delay profile $\{0.0, 0.2, 0.5, 1.6, 2.3, 5.0\}\mu s$ and power profile $\{0.189, 0.379, 0.239, 0.095, 0.061, 0.037\}$. The channel output signal-to-noise ratio (SNR) is E_b/N_0. An OFDM system with 32 subcarriers and binary phase shift keying (BPSK) is simulated. A training symbol is transmitted at the Nth subcarrier to solve the scaling ambiguity. The performance of one-tap equalization with perfect knowledge of the CIR – perfect channel state information (CSI) – provides the benchmark.

Proposed detectors are tested on OFDM systems with the above simulation parameters under different SNR over a 6-ary exponential PDP channel. Figure 5.4

shows the MSE of channel estimation which is defined as

$$\text{MSE} = E\left\{\sum_{l=1}^{L} |h_l - \hat{h}_l|^2\right\}. \tag{5.54}$$

The semi-blind detector (5.14) with the SD has MSE performance identical to that of the blind detector (5.10) with the SD. In high SNR, the semi-blind detector with V-BLAST performs close to that with the SD, while the blind detector with V-BLAST still has a 1.2-dB gap over that with the SD at MSE= 5×10^{-4}.

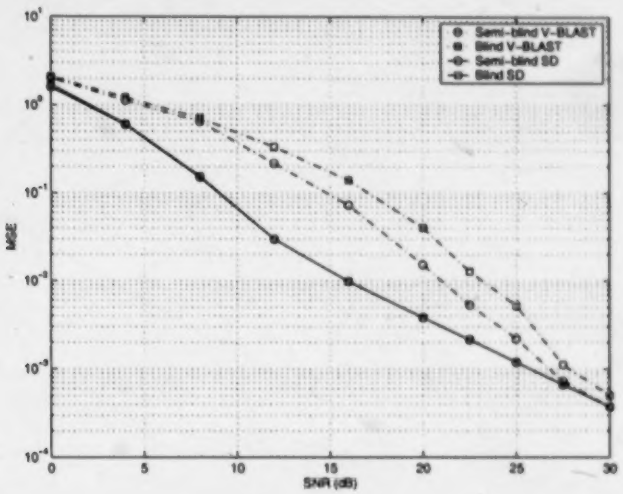

Fig. 5.4. MSE of the joint ML estimation of the channel response versus SNR for an OFDM system with $N = 32$ and BPSK in a 6-ray exponential PDP channel.

In Fig. 5.5, the BER performance of the OFDM system is compared with that of the benchmark. Both detectors with the SD are within 0.5 dB of the benchmark in high SNR. The performance of V-BLAST detection for the semi-blind detector is comparable to that of the SD in high SNR. In low SNR, the gap between SD and V-BLAST can be as large as 5 dB. The average computational complexity as a function of the SNR is given in Fig. 5.6. Note that the complexity of the exhaustive search is 1.81×10^{13} flops while even for an SNR of 10 dB all detectors' complexities are within 10^6 flops by using SD. The computational time can be saved significantly. The complexity of both detectors increases with the increase of SNR. At 0 dB, the semi-blind detector is 32 times faster than the blind detector. The blind detector

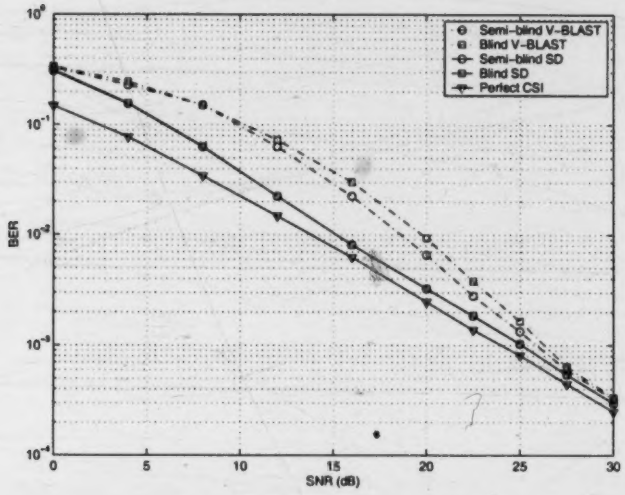

Fig. 5.5. Bit error rate of the joint ML estimation algorithm versus SNR for an OFDM system with $N = 32$ and BPSK in a 6-ray exponential PDP channel.

has higher complexity than the semi-blind detector in low SNR. This is possibly due to the inherent rank-deficiency in (5.9d) while the complexity is greatly reduced compared with the exhaustive search in the first $N - L$ variables [16] in the blind detector. When the SNR is larger than 25 dB, both the semi-blind detector and the blind detector have identical complexity. The semi-blind detector is preferable in low SNR when the channel statistics are known at the receiver.

We also compare the different algorithms over the COST 207 TU channel model, described above. The channel is assumed to be constant for 100 OFDM symbols. The channel is estimated using the first OFDM symbol, and the remaining 99 OFDM symbols are detected using the channel estimate. Fig. 5.7 shows the BER of V-BLAST and SD detection for the semi-blind detector and SD detection for the blind detector. The blind and semi-blind detectors almost achieve the bound given by one-tap equalization with perfect CIR. As shown in Fig. 5.5, the V-BLAST detection for the semi-blind detector is comparable to that of SD. This result seems to contradict the results given in [28], where great performance improvement is achieved by using the SD. The reason may be that the order of the constellation is not very high.

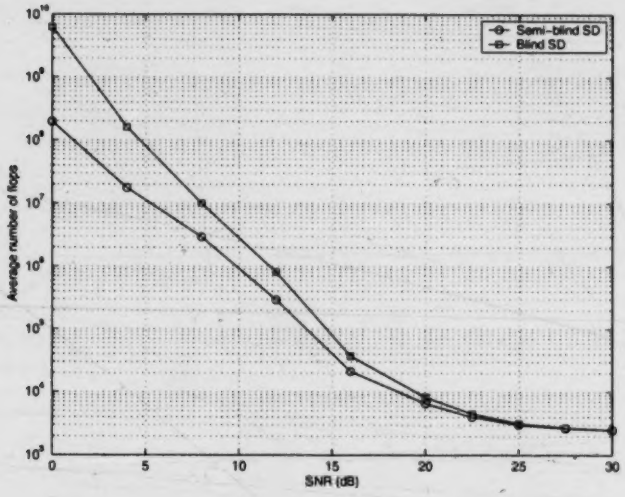

Fig. 5.6. The computational complexity versus SNR for an OFDM system with $N = 32$ and BPSK in a 6-ray exponential PDP channel.

The effect of semi-blind design mismatch is shown in Fig. 5.8. The semi-blind detector is designed for UPDP and SNR=20 dB and evaluated for a 6-ary exponentially decaying power-delay profile. The BER of the robust design is compared with perfect $\mathbf{R}_h$ and σ_n^2. From the figure, the BER performance of the two detectors are almost the same. This figure confirms the robust design criteria.

The performance of the blind suboptimal detector with SD and V-BLAST, the semi-blind suboptimal detector with SD and V-BLAST, the semi-blind optimal detector with modified SD are compared with that of the benchmark in Fig. 5.9 for an OFDM system with $N = 32$ and the 4PAM constellation. The suboptimal detectors are denoted by "Approx" in the figure. The suboptimal blind and semi-blind detectors with SD and V-BLAST perform close in high SNR. But all of them have a 4-dB performance loss at BER= 2×10^{-3}. The optimal semi-blind detector performs close to the benchmark in high SNR. At BER= 2×10^{-3}, it has only a 0.5-dB loss. In the figure, we only plot the optimal semi-blind BER curve above 25 dB. This is due to the fact that the bound for $\mathbf{x}^H\mathbf{Y}_D^H\mathbf{B}^{-1}\mathbf{Y}_D\mathbf{x}$ given by $\mathbf{x}^H\mathbf{Y}_D^H\mathbf{A}^{-1}\mathbf{Y}_D\mathbf{x}$ becomes weak in low SNR and the complexity becomes exponential.

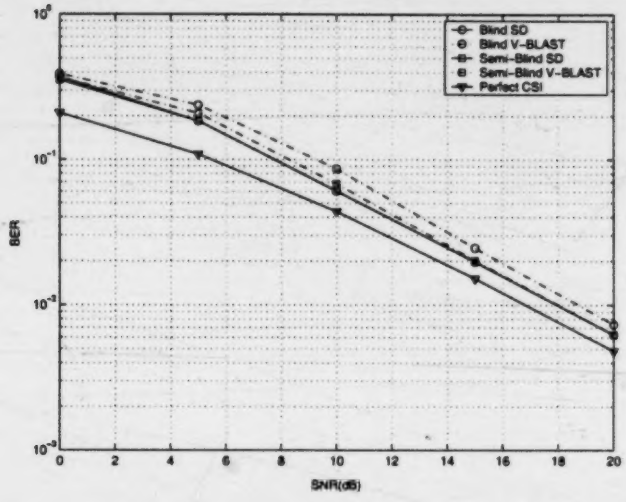

Fig. 5.7. Bit error rate of the joint ML estimation algorithm versus SNR for an OFDM system with $N = 32$ and BPSK in a TU channel

The blind detector needs knowledge of the channel length L. If L is overestimated, the effect of channel length overestimation is presented in Fig. 5.10. The simulation is performed over a 6-ary exponential PDP channel. The blind detector is evaluated at $L = 6, 8, 10$. The overestimation of L causes a performance loss in low SNR. However, in high SNR, the performance loss is negligible. At BER=4×10^{-4}, the detector with $L = 8$ has less than 0.1 dB loss over that with perfect L. When L increases to 10, the gap is still less than 0.5 dB. Therefore, our blind detector is insensitive to the overestimation of L.

Fig. 5.11 compares the performance of the enhanced data detector (5.25) and the one-tap equalization using pilot-aided channel estimates with estimation errors. A 3-ary channel is simulated. We simulate BPSK-OFDM at a data rate of 500 kbps with different number of pilots. The pilots are uniformly distributed. At low SNR, the enhanced detector performance varies slightly as the number of pilots increases from 4 to 16. At high SNR, the enhanced detector performs virtually unchanged as the number of pilots increases and it gains 1.6 dB, 0.75 dB and 0.2 dB over one-tap equalization with 4, 8 and 16 pilots, respectively. The enhanced detector performs

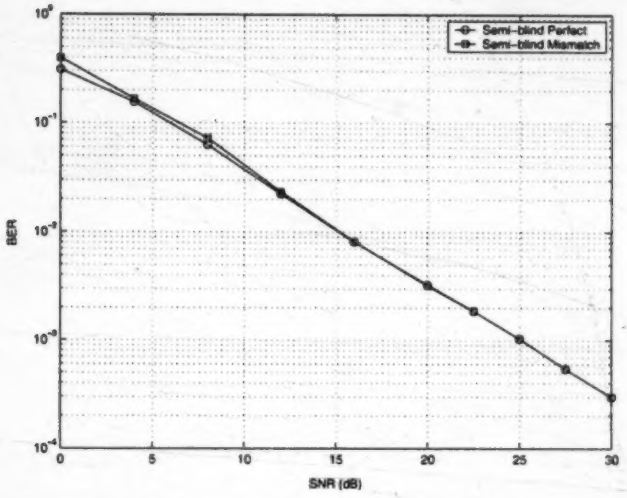

Fig. 5.8. Effects of semi-blind detector design mismatch in an OFDM system with $N = 32$ and BPSK. The channel is simulated using an exponential PDP. But in the semi-blind detector, the uniform PDP is assumed

0.4 dB within the benchmark.

5.7 Conclusion

We have developed new blind and semi-blind data detectors and channel estimators for OFDM systems. We also derived an enhanced data detector to mitigate the effect of channel estimation error. Our data detectors are maximum likelihood and require minimizing a complex, integer quadratic. The quadratic for the blind detector suffers from rank deficiency, to which we gave an efficient solution. Since the semi-blind detector uses both channel correlation and noise level, we have presented noise-variance and power-delay-profile estimators using the cyclic prefix in each OFDM block. We have also provided simple adaptations of the SD algorithm to handle M-PSK constellations and to achieve reduced complexity. We considered how the semi-blind detector performs under residual mismatch and generalized the basic data detectors

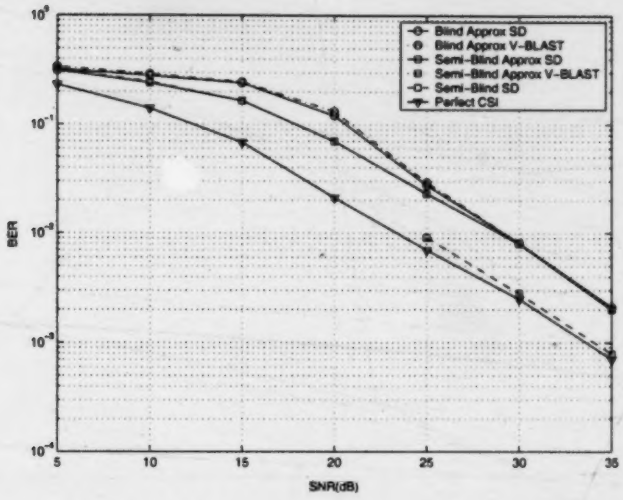

Fig. 5.9. Bit error rate of the joint ML estimation algorithm versus SNR for an OFDM system with $N = 32$ and 4PAM in a 3-ray exponential PDP channel.

to non-unitary constellations. Simulation results show that the proposed detectors perform close to the ideal case. They may also be extended to MIMO-OFDM systems and OFDM over fast fading channels. These applications are currently being investigated.

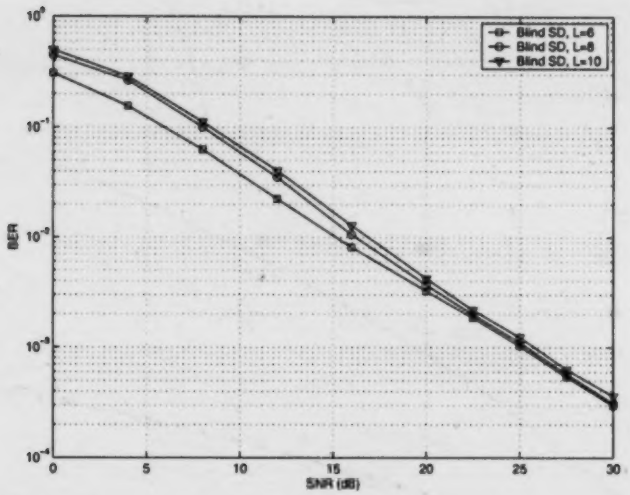

Fig. 5.10. The BER versus SNR for an BPSK OFDM system with $N = 32$ and assuming different channel length L.

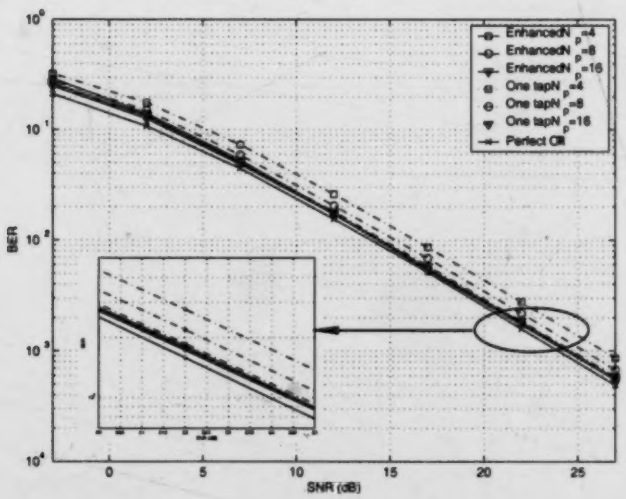

Fig. 5.11. The performance of the enhanced data detector as a function of the number of pilots for a BPSK-OFDM system with $N = 32$.

Chapter 6

Conclusion

With the rapid growth of the wireless industry, there is a demand for low-complexity receivers. Large potential for the complexity reduction of high-performance VLSI signal processing circuits motivates the development of the efficient detection algorithms. In this thesis, we have studied efficient detectors for MIMO and OFDM systems, which is critical to achieve practicable solutions for next-generation wireless communication systems.

In Chapter 2, we developed a unified framework for efficient data detection of spatial multiplexing MIMO systems, which includes the well-known algorithms such as ZF-BLAST [1], SD [2], combined ML and ZF-DFD [18] and the B-Chase detector [19] as special cases. This framework provides a performance-complexity tradeoff. We also considered the relaxation approach to the MIMO detection and presented a class of constrained linear detectors and a class of constrained decision feedback detectors. Moreover, a polynomial constrained detector was proposed and solved using the penalty function approach and differential equations. Multistage sphere decoder for high order constellation applications was also derived, which exploits the fact that many higher-order signal constellations can naturally be decomposed into several lower-order constellations.

In Chapter 3, we derived an ML decision metric for blind decoding of OSTBC in a quasi static fading channel by using the linear dispersion property of OSTBC. The decision metric results in a quadratic minimization problem, which can be solved using V-BLAST, SD or the detectors developed in Chapter 2. To remove the need of pilots, novel approaches for totally blind decoding were presented using two different PSK constellations and a superimposed training scheme. We also gave an MMSE channel estimator and derive the CRB. Power allocation issues were also discussed.

In Chapter 4, we developed efficient detectors for DUSTM, where the detection problem was formulated as a one-dimensional NP-hard problem. The BID was derived for single symbol detection with diagonal constellations using the extended Euclidean algorithm [21], well-known for determining the greatest common divisor (gcd) of two integers. Our BID achieves significant computational savings over the ML search, especially in high SNR. We have also developed four BID variants for multiple symbol detection of DUSTM. The first two were ML and used BnB, the third one was suboptimal, which first uses BID to generate a candidate subset and then exhaustively searches over the reduced space, and the last one generalized decision-feedback differential detection.

We developed new blind and semi-blind data detectors for OFDM systems in Chapter 5, which also result in a quadratic form in data symbols. We proposed a cyclic-prefix based channel correlation and noise variance estimation algorithm for the semi-blind detector, which requires both channel correlation and noise variance. To mitigate the effect of channel estimation errors in pilot aided LS channel estimator, an enhanced data detector was also derived by noting that the true CIR can be modeled as complex Gaussian with mean being the LS channel estimate itself. The LS channel estimate thus gives a prior on the true channel and averaging the likelihood function over the prior distribution gives the enhanced detector.

Efficient receiver design is a hot research topic and has attracted great interest in the wireless communications community. As such technology is making its way

from the research labs into industry standards, it will play a crucial role in designing next-generation wireless communications systems as well as reengineering existing systems to obtain higher bandwidths. Therefore, we expect to see increasing interest in this area. The study in this thesis only scratches the tip of the iceberg and many important problems remain to be answered:

- In this thesis, we only discussed efficient data detection for uncoded systems. However, in many practical systems, an error-correcting code is usually employed. The detectors proposed in this thesis may also be extended to coded systems.
- While several relaxation approaches have been proposed in the literature, all of them are not tight enough. Tight relaxation still remains an open problem in both communications and optimization theory.
- In practical systems, super-orthogonal block codes can achieve higher data rates [92]. It would be interesting to derive blind decoders for such codes.
- Our blind and semiblind data detectors for OFDM systems may also be extended to MIMO-OFDM systems. Also, OFDM transmission over doubly selective channels is a challenging problem. Low-complexity channel estimators and equalizers may be derived for transmission over doubly selective channels.

References

[1] G. D. Golden, G. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using the V-BLAST space-time communication architecture," *Electronics Letters*, vol. 35, no. 1, pp. 14–15, Jan. 1999.

[2] E. Viterbo and J. Bouros, "A universal lattice code decoder for fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1639–1642, Jul. 1999.

[3] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806 – 2818, Aug. 2005.

[4] A. Burg, M. Borgmann, C. Simon, M. Wenk, M. Zellweger, and W. Fichtner, "Performance tradeoffs in the VLSI implementation of the sphere decoding algorithm," in *Proc. IEE 3G Mobile Communication Technologies Conference*, London, UK, Oct. 2004.

[5] D. Garrett, L. Davis, S. ten Brink, B. Hochwald, and G. Knagge, "Silicon complexity for maximum likelihood MIMO detection using spherical decoding," *IEEE J. Solid-State Circuits*, vol. 39, no. 9, pp. 1544 – 1552, Sept. 2004.

[6] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Processing*, vol. 53, no. 4, pp. 1474 – 1484, April 2005.

[7] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1456 – 1467, July 1999.

[8] V. Tarokh and H. Jafarkhani, "A differential detection scheme for transmit diversity," *IEEE J. Select. Areas Commun.*, vol. 18, no. 7, pp. 1169 – 1174, July 2000.

[9] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2041 – 2052, Dec. 2000.

[10] B. L. Hughes, "Differential space-time modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2567 – 2578, Nov. 2000.

[11] J. H. Conway and N. J. Sloane, *Sphere packing, Lattices and Groups*, 3rd ed. New York: Springer-Verlag, 1998.

[12] G. Ginis and J. M. Cioffi, "On the relation between V-BLAST and the GDFE," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 364 – 366, Sept. 2001.

[13] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, inlcuding a complexity analysis," *Math. Computation*, vol. 44, pp. 463–471, Apr. 1985.

[14] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Programming*, vol. 66, pp. 181–191, 1994.

[15] L. Brünel, "Optimum multiuser detection for MC-CDMA systems using sphere decoding," in *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, Sept. 2001, pp. A–16 – A–20.

[16] M. O. Damen, A. Chkeif, and J. C. Belfiore, "Lattice code decoder for space-time codes," *IEEE Commun. Lett.*, vol. 4, no. 5, pp. 161 – 163, May 2000.

[17] H. Vikalo and B. Hassibi, "On maximum-likelihood sequence detection for multiple antenna systems over dispersive channels," *EURASIP Journal on Applied Signal Processing*, vol. Special Issue on Space-Time Coding, pp. 525–531, May 2002.

[18] W.-J. Choi, R. Negi, and J. M. Cioffi, "Combined ML and DFE decoding for the V-BLAST system," in *Proc. of ICC*, vol. 3, June 2000, pp. 1243 – 1248.

[19] D. W. Waters and J. R. Barry, "The chase family of detection algorithms for multiple-input multiple-output channels," in *Proc. of GLOBECOM*, vol. 4, Nov. 2004, pp. 2635 – 2639.

[20] J. A. Heller, "Feedback decoding of convolutional codes," in *Advances in Commun. Syst.*, A. J. Viterbi, Ed., vol. 4, NY: Academic Press, 1975, pp. 261–278.

[21] E. Bach and J. Shallit, *Algorithmic Number Theory, Vol. 1: Efficient Algorithms.* Cambridge, MA: MIT Press, 1996.

[22] A. Benjebbour, H. Murata, and S. Yoshida, "Comparison of ordered successive receivers for space-time transmission," in *Proc. of VTC Fall*, vol. 4, Oct. 2001, pp. 2053 – 2057.

[23] A. Yener, R. D. Yates, and S. Ulukus, "CDMA multiuser detection: a nonlinear programming approach," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1016 – 1024, June 2002.

[24] S. Thoen, L. Deneire, L. V. der Perre, M. Engels, and H. D. Man, "Constrained least squares detector for OFDM/SDMA-based wireless networks," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 129 – 140, Jan. 2003.

[25] W.-K. Ma, T. Davidson, K. M. Wong, Z.-Q. Luo, and P.-C. Ching, "Quasi-maximum-likelihood multiuser detection using semi-definite relaxation with application to synchronous CDMA," *IEEE Trans. Signal Processing*, vol. 50, no. 4, pp. 912 – 922, April 2002.

[26] W.-K. Ma, P.-C. Ching, and Z. Ding, "Semidefinite relaxation based multiuser detection for M-ary PSK multiuser systems," *IEEE Trans. Signal Processing*, vol. 52, no. 10, pp. 2862 – 2872, Oct. 2004.

[27] B. Tarokh and H. R. Sadjadpour, "Construction of OFDM M-QAM sequences with low peak-to-average power ratio," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 25 – 28, Jan. 2003.

[28] M. O. Damen, E. Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[29] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill Education, 2001.

[30] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[31] X. Li, H. Huang, G. Foschini, and R. Valenzuela, "Effects of iterative detection and decoding on the performance of BLAST," in *Proc. of GLOBECOM*, vol. 2, Nov 2000, pp. 1061 – 1066.

[32] A. Papoulis and S. Pillai, *Probability, Random Variables and Stochastic Processes.* McGraw-Hill Education, Dec. 2001.

[33] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, reprint ed. Waveland Press, 1990.

[34] S. S. Skiena, *The Algorithm Design Manual*, 1st ed. Springer-Verlag, 1997.

[35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 2002.

[36] D. Chase, "Class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 170 – 182, Jan. 1972.

[37] R. Fischer and C. Windpassinger, "Real versus complex-valued equalisation in V-BLAST systems," *Electronics Letters*, vol. 39, no. 5, 6, pp. 470 – 471, March 2003.

[38] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. New York: Academic Press, 1982.

[39] F. Aluffi-Pentini, V. Parisi, and F. Zirilli, "A differential-equations algorithm for nonlinear equations," *ACM Trans. on Mathematical Software*, vol. 10, no. 3, pp. 299 – 316, Sept. 1984.

[40] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1451 – 1458, Oct. 1998.

[41] P. Stoica and G. Ganesan, "Space-time block codes: trained, blind and semi-blind detection," in *Proc. of ICASSP*, May 2002, pp. 7–11.

[42] A. L. Swindlehurst and G. Leus, "Blind and semi-blind equalization for generalized space-time block codes," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2489 – 2498, Oct. 2002.

[43] W.-K. Ma, P. C. Ching, T. N. Davidson, and X. G. Xia, "Blind maximum-likelihood decoding for orthogonal space-time block codes: a semidefinite relaxation approach," in *Proc. of GLOBECOM*, Dec. 2003, pp. 2094 – 2098.

[44] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804 – 1824, July 2002.

[45] J. W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.

[46] M. C. Necker and G. L. Stuber, "Totally blind channel estimation for OFDM on fast varying mobile radio channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1514 – 1525, Sept. 2004.

[47] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2515 – 2528, Oct. 2002.

[48] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 38, no. 3, pp. 300 – 308, March 1990.

[49] B. Li, "A new reduced-complexity algorithm for multiple-symbol differential detection," *IEEE Commun. Lett.*, vol. 7, no. 6, pp. 269 – 271, June 2003.

[50] K. M. Mackenthun, "A fast algorithm for multiple-symbol differential detection of MPSK," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1471 – 1474, Feb./Mar./Apr. 1994.

[51] I. Motedayen-Aval and A. Anastasopoulos, "Polynomial-complexity noncoherent symbol-by-symbol detection with application to adaptive iterative decoding of turbo-like codes," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 197 – 207, Feb. 2003.

[52] L. Lampe, R. Schober, V. Pauli, and C. Windpassinger, "Multiple-symbol differential sphere decoding," in *Proc. of ICC*, June 2004, pp. 787 – 791.

[53] P. Kam and C. Teh, "Reception of PSK signals over fading channels via quadrature amplitude estimation," *IEEE Trans. Commun.*, vol. 31, no. 8, pp. 1024 – 1027, Aug. 1983.

[54] R. Schober, W. H. Gerstacker, and J. B. Huber, "Decision-feedback differential detection of MDPSK for flat Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1025 – 1035, July 1999.

[55] R. Schober and L. H. J. Lampe, "Noncoherent receivers for differential space-time modulation," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 768 – 777, May 2002.

[56] B. Bhukania and P. Schniter, "On the robustness of decision-feedback detection of DPSK and differential unitary space-time modulation in Rayleigh-fading

channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1481–1489, Sept. 2004.

[57] E. Chiavaccini and G. M. Vitetta, "Further results on differential space-time modulations," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1093 – 1101, July 2003.

[58] K. L. Clarkson, W. Sweldens, and A. Zheng, "Fast multiple-antenna differential decoding," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 253 – 261, Feb. 2001.

[59] A. Lenstra, H. Lenstra, and L. Lovasz, "Factoring polynomials with rational coefficients," *Math. Ann.*, vol. 261, pp. 515–534, 1982.

[60] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619 – 2692, Oct. 1998.

[61] R. U. Nabar, H. Bolcskei, and A. J. Paulraj, "Outage properties of space-time block codes in correlated Rayleigh or Rician fading environments," in *Proc. of ICASSP*, May 2002, pp. 13–17.

[62] I. Dinur, G. Kindler, and S. Safra, "Approximating-CVP to within almost-polynomial factors is NP-hard," in *Proc. of IEEE Symposium on Foundations of Computer Science*, 1998.

[63] M. Lempel and A. Paz, "An algorithm for finding a shortest vector in a two dimensional modular lattice," *Theoretical Computer Science*, vol. 125, pp. 229–241, 1994.

[64] G. Rote, "Finding a shortest vector in a two-dimensional lattice modulo m," *Theoretical Computer Science*, vol. 172, pp. 303–308, 1997.

[65] C. Ling, W. H. Mow, K. H. Li, and A. C. Kot, "Multiple-antenna differential lattice decoding," to appear in *IEEE J. Select. Areas Commun.*

[66] G. Havas, "On the complexity of the extended Euclidean algorithm," in *Electronic Notes in Theoretical Computer Science*, J. Harland, Ed. Elsevier, 2003.

[67] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-state sequence estimation with set partitioning and decision feedback," *IEEE Trans. Commun.*, vol. 36, no. 1, pp. 13 – 20, Jan. 1988.

[68] A. Duel-Hallen and C. Heegard, "Delayed decision-feedback sequence estimation," *IEEE Trans. Commun.*, vol. 37, no. 5, pp. 428 – 436, May 1989.

[69] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge; New York: Cambridge University Press, 1985.

[70] P. Soma, D. S. Baum, V. Erceg, R. Krishnamoorthy, and A. J. Paulraj, "Analysis and modeling of multiple-input multiple-output (MIMO) radio channel based on outdoor measurements conducted at 2.5 GHz for fixed BWA applications," in *Proc. of ICC*, May 2002, pp. 272 – 276.

[71] G. L. Stuber, J. R. Barry, S. W. Mclaughlin, Y. G. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271 – 294, Feb. 2004.

[72] O. Edfors, M. Sandell, J.-J. van de Beek, S. Wilson, and P. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[73] Y. Li, J. Cimini, L.J., and N. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902 – 915, July 1998.

[74] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.

[75] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1699 – 1712, July 2002.

[76] R. Heath and G. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Processing*, vol. 47, no. 3, pp. 848–856, Mar. 1999.

[77] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141–150, Jan. 2003.

[78] N. Chotikakamthorn and H. Suzuki, "On identifiability of OFDM blind channel estimation," in *Proc. of VTC Fall*, vol. 4, Sep. 1999, pp. 2358–2361.

[79] Y. Song, S. Roy, and L. Akers, "Joint blind estimation of channel and data symbols in OFDM," in *Proc. of VTC Spring*, vol. 1, May. 2000, pp. 46–50.

[80] M. Luise, R. Reggiannini, and G. M. Vitetta, "Blind equalization/detection for OFDM signals over frequency-selective channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1568 – 1578, Oct. 1998.

[81] M.-X. Chang and Y. T. Su, "Blind and semiblind detections of OFDM signals in fading channels," *IEEE Trans. Commun.*, vol. 52, no. 5, pp. 744 – 754, May 2004.

[82] D. Warrier and U. Madhow, "Spectrally efficient noncoherent communication," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 651 – 668, March 2002.

[83] P. Stoica, H. Vikalo, and B. Hassibi, "Joint maximum-likelihood channel estimation and signal detection for SIMO channels," in *Proc. of ICASSP*, vol. 4, April 2003, pp. 13–16.

[84] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer in the 5 GHz band," in *IEEE Std 802.11a*, Sept. 1999.

[85] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*. New York: McGraw-Hill, 1966.

[86] S. He and M. Torkelson, "Effective SNR estimation in OFDM system simulation," in *Proc. of GLOBECOM*, Nov. 1998, pp. 945 – 950.

[87] X. Xu, Y. Jing, and X. Yu, "Subspace-based noise variance and SNR estimation for OFDM systems," in *Proc. of WCNC*, Mar. 2005, pp. 23 – 26.

[88] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467 – 479, Mar. 2001.

[89] S. Boumard, "Novel noise variance and SNR estimation algorithm for wireless MIMO OFDM systems," in *Proc. of GLOBECOM*, Dec. 2003, pp. 1330 – 1334.

[90] J. J. van de Beek, M. Sandell, and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, no. 7, pp. 1800–1805, Jul. 1997.

[91] G. L. Stuber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer Academic, 2001.

[92] H. Jafarkhani, "A quasi-orthogonal space-time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1 – 4, Jan. 2001.